



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10175884 A**(43) Date of publication of application: **30 . 06 . 98**

(51) Int. Cl.

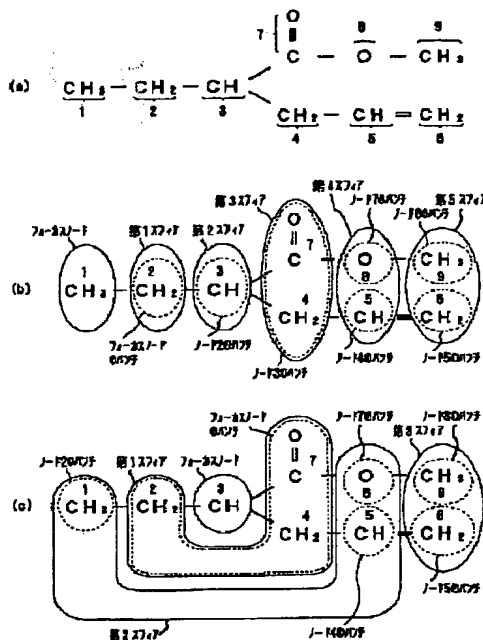
C07B 61/00
G06F 17/30(21) Application number: **08352326**(22) Date of filing: **12 . 12 . 96**(71) Applicant: **SUMITOMO CHEM CO LTD**(72) Inventor: **MASUI HIDEYUKI**
OKUMA NOBUYUKI(54) **CODING OF PARTIAL STRUCTURE OF ORGANIC COMPOUND AND MEDIUM RECODING PROGRAM FOR CODING PARTIAL STRUCTURE**

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a method for coding the partial structure of an organic compound, enabling to more precisely express the environments of elements to which hydrogen atoms are bound than by other methods.

SOLUTION: This method for coding the partial structure of an organic compound comprises defining one node 1 of the organic compound comprising plural nodes 1-9 as a focus node and subsequently coding the partial structure ranging from the focus node 1 to the sphere N with focus information describing the node code of the focus node 1 and sphere information describing the node codes and bond codes of the nodes contained in the spheres 1 to N. Therein, node codes different in response to the number of bound hydrogen atoms are assigned to the nodes of carbon, nitrogen and oxygen, respectively, and only one kind of node code is assigned to the nodes of other element except the elements, respectively.

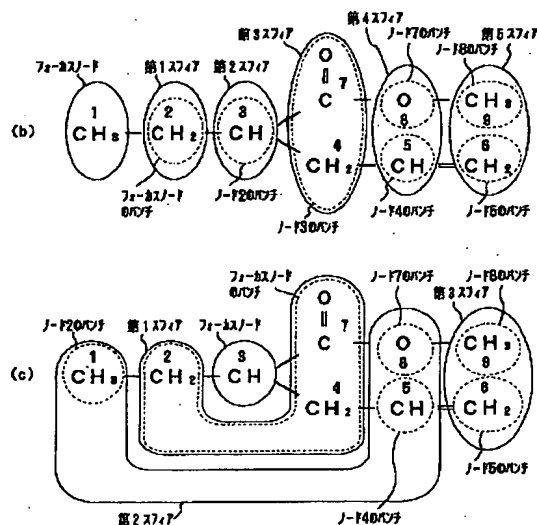
COPYRIGHT: (C)1998,JPO



(11)特許出願公開番号

(43)公開日 平成10年(1998)6月30日

370 F



【特許請求の範囲】

【請求項1】 複数のノードから構成される有機化合物の1つのノードをフォーカスノードとし、第Nスフィアまでの部分構造を、フォーカスノードのノードコードを記述したフォーカス情報と、各スフィアに含まれるノードのノードコードおよびボンドコードを記述したN個のスフィア情報とによって、コード化する方法であって、少なくとも炭素のノードに対してはその結合水素数に応じて異なるノードコードを割り当てるようにしたことを特徴とする有機化合物の部分構造のコード化方法。

【請求項2】 炭素に加え、窒素および酸素のノードに対してもその結合水素数に応じて異なるノードコードを割り当て、それ以外の元素のノードに対しては1種類のノードコードを割り当てるようにしたことを特徴とする請求項1記載の有機化合物の部分構造のコード化方法。

【請求項3】 フォーカス情報、第1スフィア情報、…、第Nスフィア情報の順にコード化していく際に、或るノードXに直接結合するノードYをコード化するとき該コード化位置以前に既にノードYをコード化していた場合には、ノードYのノードコードの代わりに閉環コードを使用し、該コード化位置以降にノードYに直接結合するノードとしてノードXが現れたときに該ノードXのノードコードの代わりに前記ノードYに使用した閉環コードと同じ閉環コードを使用し、且つ、閉環コードで表現されたノードについてはそれを起点とする以降のノードのコード化を省略するようにしたことを特徴とする請求項1または2記載の有機化合物の部分構造のコード化方法。

【請求項4】 1つの部分構造のコード化において複数の閉環コードを使用する必要がある場合、予め定められたそれぞれ異なる複数の閉環コードから循環的に1つずつ閉環コードを選択して使用することを特徴とする請求項3記載の有機化合物の部分構造のコード化方法。

【請求項5】 フォーカス情報と第1スフィア情報との間に第1の区切り記号コードを置き、各スフィア情報間に第2の区切り記号コードを置き、各スフィア内のパンチ情報間に第3の区切り記号コードを置くようにし、更に、パンチ内においては所定の優先順位に従ってノードコードとボンドコードとの組を配列するようにしたことを特徴とする請求項1、2、3または4記載の有機化合物の部分構造のコード化方法。

【請求項6】 有機化合物を構成する元素の種類および各元素間の結合状態を記述した入力ファイルを読み込む第1ステップと、有機化合物を構成する元素のうちの1つの元素をフォーカスノードに定める第2ステップと、少なくとも炭素のノードに対してはその結合水素数に応じて異なるノードコードを割り当てるようにしつつ、フォーカスノードのノードコードを記述したフォーカス情報をルートとし、第1～第Nスフィアに含まれるノード

のノードコードおよびボンドコードを記述したN個のスフィア情報をそれぞれ第1～第N階層とする木構造を作成する第3ステップと、

木構造中のスフィア情報における各パンチ内のノードコードとボンドコードとの組の並びが所定の優先順位に沿うように木構造を正規化する第4ステップと、正規化された木構造をフォーカス情報、第1スフィア情報、…、第Nスフィア情報の順に走査して、フォーカス情報と第1スフィア情報との間に第1の区切り記号コードを置き、各スフィア情報間に第2の区切り記号コードを置き、各スフィア内のパンチ情報間に第3の区切り記号コードを置いたコードを生成する第5ステップと、生成したコードを出力ファイルに出力する第6ステップとを含むことを特徴とする有機化合物の部分構造のコード化方法。

【請求項7】 第3ステップにおいて、炭素に加え、窒素および酸素のノードに対してもその結合水素数に応じて異なるノードコードを割り当て、それ以外の元素のノードに対しては1種類のノードコードを割り当てるようにしたことを特徴とする請求項6記載の有機化合物の部分構造のコード化方法。

【請求項8】 第3ステップにおいて、フォーカス情報、第1スフィア情報、…、第Nスフィア情報の順にコード化していく際に、或るノードXに直接結合するノードYをコード化するとき該コード化位置以前に既にノードYをコード化していた場合には、ノードYのノードコードの代わりに閉環コードを使用し、該コード化位置以降にノードYに直接結合するノードとしてノードXが現れたときに該ノードXのノードコードの代わりに前記ノードYに使用した閉環コードと同じ閉環コードを使用し、且つ、閉環コードで表現されたノードについてはそれを起点とする以降のノードのコード化を省略するようにしたことを特徴とする請求項6または7記載の有機化合物の部分構造のコード化方法。

【請求項9】 第3ステップにおいて、1つの部分構造のコード化において複数の閉環コードを使用する必要がある場合、予め定められたそれぞれ異なる複数の閉環コードから循環的に1つずつ閉環コードを選択して使用することを特徴とする請求項8記載の有機化合物の部分構造のコード化方法。

【請求項10】 第2ステップにおいて、入力ファイルにケミカルシフト値が記述されている元素をフォーカスノードとし、且つ、第6ステップにおいて、生成した部分構造のコードと共にそのケミカルシフト値を出力ファイルに出力することを特徴とする請求項6、7、8または9記載の有機化合物の部分構造のコード化方法。

【請求項11】 コンピュータに下記のステップを実行させるための部分構造コード化プログラムを記録した媒体。

10

20

30

40

50

第1ステップ;有機化合物を構成する元素の種類および各元素間の結合状態を記述した入力ファイルを読み込む。

第2ステップ;有機化合物を構成する元素のうちの1つの元素をフォーカスノードに定める。

第3ステップ;少なくとも炭素のノードに対してはその結合水素数に応じて異なるノードコードを割り当てるようにしつつ、フォーカスノードのノードコードを記述したフォーカス情報をルートとし、第1～第Nスフィアに含まれるノードのノードコードおよびボンドコードを記述したN個のスフィア情報をそれぞれ第1～第N階層とする木構造を作成する。

第4ステップ;木構造中のスフィア情報における各パンチ内のノードコードとボンドコードとの組の並びが所定の優先順位に沿うように木構造を正規化する。

第5ステップ;正規化された木構造をフォーカス情報、第1スフィア情報、…、第Nスフィア情報の順に走査して、フォーカス情報と第1スフィア情報との間に第1の区切り記号コードを置き、各スフィア情報間に第2の区切り記号コードを置き、各スフィア内のパンチ情報間に第3の区切り記号コードを置いたコードを生成する。

第6ステップ;生成したコードを出力ファイルに出力する。

【請求項12】 第3ステップが、炭素に加え、窒素および酸素のノードに対してもその結合水素数に応じて異なるノードコードを割り当て、それ以外の元素のノードに対しては1種類のノードコードを割り当てる手順を含むことを特徴とする請求項11記載の部分構造コード化プログラムを記録した媒体。

【請求項13】 第3ステップが、フォーカス情報、第1スフィア情報、…、第Nスフィア情報の順に木構造を作成していく際に、或るノードXに直接結合するノードYをコードするとき既にノードYをコード化していた場合には、ノードYのノードコードの代わりに閉環コードを使用し、以降にノードYに直接結合するノードとしてノードXが現れたときに該ノードXのノードコードの代わりに前記ノードYに使用した閉環コードと同じ閉環コードを使用し、且つ、閉環コードで表現されたノードについてはそれを起点とする以降のノードのコード化を省略する手順を含むことを特徴とする請求項11または12記載の部分構造コード化プログラムを記録した媒体。

【請求項14】 第3ステップが、1つの部分構造のコード化において複数の閉環コードを使用する必要がある場合、予め定められたそれぞれ異なる複数の閉環コードから循環的に1つずつ閉環コードを選択して使用する手順を含むことを特徴とする請求項13記載の部分構造コード化プログラムを記録した媒体。

【請求項15】 前記第2ステップが、入力ファイルにケミカルシフト値が記述されている元素をフォーカスノードとする手順を含み、

前記第6ステップが、生成した部分構造のコードと共にそのケミカルシフト値を出力ファイルに出力する手順を含むことを特徴とする請求項11, 12, 13または14記載の部分構造コード化プログラムを記録した媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は有機化合物の部分構造のコード化方法に関し、特に水素の結合する元素の環境をより精度良く表現できるコード化方法に関する。

【0002】

【従来の技術】未知の有機化合物のスペクトルを測定し、コンピュータによってその構造を解析、確認する構造解析システムが従来より種々提案されている(たとえば、コンピュータ・ケミストリー・シリーズ1、『CHEMICS-コンピュータによる構造解析-』, 佐々木慎一, 船津公人著, 1994年6月1日, 共立出版株式会社発行)。

【0003】このような有機化合物の構造解析システムでは、構造の明らかな有機化合物の部分構造とスペクトルデータとの対応を知識ベースとして予め蓄積しておき、未知の有機化合物から得られたスペクトルデータによって知識ベースを検索し、未知の有機化合物に含まれる可能性のある部分構造を推定する。そして、この推定された部分構造を基にして全体構造を組み立て、最後に、この組み立てられた全体構造を検定し、全体構造を特定する。

【0004】このため、この種のシステムでは、有機化合物の部分構造を如何にしてコンピュータによって処理可能なコードにコード化するかが重要になってくる。

【0005】コンピュータで処理可能な有機化合物の構造表記方式としては、WLN線形表記、結合表(Con-
e-
c-
t-
i-
o-
n
T-
a-
b-
l-
e), HOSE Codeなど、数多くの表記方式が知られている。しかし、WLN線形表記、結合表は有機化合物の全体構造の表記には適するが、部分構造の表記方法としては適切でなく、使用できない。このため、知識ベースのための部分構造表記方式としては、HOSE Codeが良く使用される。これはBASF社(独)で開発された、有機化合物の炭素原子に着目した表記方式で、基本的に ^{13}C -NMRスペクトルを対象として開発されたものである。

【0006】HOSE Codeでは、部分構造を記述する場合に、中心となる元素をフォーカスとし、このフォーカスを中心として第1スフィア、第2スフィア、…等のスフィアを定め、各スフィアに存在する元素を、炭素ならC、窒素ならN、酸素ならOというように各元素に1対1に対応するコードにコード化していく。例えば、全体構造が図55(a)に示すような構造式で表される有機化合物において、番号1の CH_3 をフォーカスとした場合、各スフィアは同図に示すものとなり、第4スフィアまでの部分構造は、HOSE Codeでは同

図(b)に示すように表記される。ここで、先頭のCは番号2のCH₂に、2番目のCは番号3のCHに、3番目のCは直後の\$と共に番号7のC=Oに、4番目のCは番号4のCH₂に、5番目のOは番号8のOに、そして最後のCは番号5のCHに、それぞれ対応している。また、「(」、「/」、「」は区切り記号である。

【0007】

【発明が解決しようとする課題】HOSE Codeは、基本的に¹³C-NMRスペクトルを対象としているため、炭素の結合する元素の環境は良く表現できる。しかし、水素の結合する元素の環境は、精度良く表現することができない。その原因は、HOSE Codeでは、各元素コードの水素数はその結合次数の合計と結合可能数との差を計算する必要があり、最終スフィアではそれ以降の結合状態が不明なため水素数の計算が不可能だからである。つまり、図55(b)のHOSE Codeの場合、最後のCに結合している水素数が1であることは分からない。

【0008】このようにHOSE Codeでは最終スフィアの水素数が不明となるため、例えば図55(a)の番号5, 6のCH, CH₂を同図(c)のように番号10, 11, 12のC, CH₂, CH₃に置き換えた別の有機化合物に対して、番号1のCH₃をフォーカスとした第4スフィアまでの部分構造を考えた場合、HOSE Codeでは同図(b)と全く同じ表記になる。これは、水素の結合する元素の環境が相違する複数の有機化合物に対して、HOSE Codeで部分構造を表記した場合、区別できないことを意味する。勿論、スフィア数を増加すれば区別できるが、1つの部分構造当たりのコード量が増え、コンピュータのメモリ資源を圧迫することになる。

【0009】このようにHOSE Codeでは、水素の結合する元素の環境を精度良く表現できないため、特に¹H-NMRスペクトル、H-H COSYスペクトル、C-H COSYスペクトル、HMBCスペクトル、HMQCスペクトル等のスペクトルを用いた有機化合物の構造解析システムの開発にあつては、水素の結合する元素の環境をより精度良く表現できるコード化方法の開発が最重要課題の一つになっている。

【0010】本発明はこのような事情に鑑みて提案されたものであり、その目的は、水素の結合する元素の環境をより精度良く表現できる有機化合物の部分構造のコード化方法を提供することにある。

【0011】

【課題を解決するための手段】本発明は、有機化合物を構成する元素をそれぞれノードとした場合に、そのような複数のノードから構成される有機化合物の1つのノードをフォーカスノードとし、第Nスフィアまでの部分構造を、フォーカスノードのノードコードを記述したフォーカス情報と、各スフィアに含まれるノードのノードコ

ードおよびボンドコードを記述したN個のスフィア情報とによって、コード化する方法であつて、少なくとも炭素のノードに対してはその結合水素数に応じて異なるノードコードを割り当てるようにしたことを特徴とする。

【0012】また、本発明の有機化合物の部分構造の他のコード化方法は、炭素に加え、窒素および酸素のノードに対してもその結合水素数に応じて異なるノードコードを割り当て、それ以外の元素のノードに対しては1種類のノードコードを割り当てるようにしたことを特徴とする。

【0013】また、本発明の有機化合物の部分構造のコード化方法は、フォーカス情報、第1スフィア情報、…、第Nスフィア情報の順にコード化していく際に、或るノードXに直接結合するノードYをコード化するとき該コード化位置以前に既にノードYをコード化していた場合には、ノードYのノードコードの代わりに閉環コードを使用し、該コード化位置以降にノードYに直接結合するノードとしてノードXが現れたときに該ノードXのノードコードの代わりに前記ノードYに使用した閉環コードと同じ閉環コードを使用し、且つ、閉環コードで表現されたノードについてはそれを起点とする以降のノードのコード化を省略するようにしたことを特徴とする。

【0014】また、本発明の有機化合物の部分構造のコード化方法は、1つの部分構造のコード化において複数の閉環コードを使用する必要がある場合、予め定められたそれぞれ異なる複数の閉環コードから循環的に1つずつ閉環コードを選択して使用することを特徴とする。

【0015】また、本発明の有機化合物の部分構造のコード化方法は、フォーカス情報と第1スフィア情報との間に第1の区切り記号コードを置き、各スフィア情報間に第2の区切り記号コードを置き、各スフィア内のパンチ情報間に第3の区切り記号コードを置くようにし、更に、パンチ内においては所定の優先順位に従ってノードコードとボンドコードとの組を配列するようにしたことを特徴とする。

【0016】また、本発明の有機化合物の部分構造のコード化方法は、有機化合物を構成する元素の種類および各元素間の結合状態を記述した入力ファイルを読み込む第1ステップと、有機化合物を構成する元素のうちの1つの元素をフォーカスノードに定める第2ステップと、少なくとも炭素のノードに対してはその結合水素数に応じて異なるノードコードを割り当てるようにしつつ、フォーカスノードのノードコードを記述したフォーカス情報をルートとし、第1～第Nスフィアに含まれるノードのノードコードおよびボンドコードを記述したN個のスフィア情報をそれぞれ第1～第N階層とする木構造を作成する第3ステップと、木構造中のスフィア情報における各パンチ内のノードコードとボンドコードとの組の並びが所定の優先順位に沿うように木構造を正規化する第4ステップと、正規化された木構造をフォーカス情報、

10

20

30

40

50

第1スフィア情報, ..., 第Nスフィア情報の順に走査して、フォーカス情報と第1スフィア情報との間に第1の区切り記号コードを置き、各スフィア情報間に第2の区切り記号コードを置き、各スフィア内のバンチ情報間に第3の区切り記号コードを置いたコードを生成する第5ステップと、生成したコードを出力ファイルに出力する第6ステップとを含むことを特徴とする。

【0017】また、本発明の有機化合物の部分構造の他のコード化方法は、前記第3ステップにおいて、炭素に加え、窒素および酸素のノードに対してもその結合水素数に応じて異なるノードコードを割り当て、それ以外の元素のノードに対しては1種類のノードコードを割り当てるようにしたことを特徴とする。

【0018】また、本発明の有機化合物の部分構造のコード化方法は、前記第3ステップにおいて、フォーカス情報、第1スフィア情報, ..., 第Nスフィア情報の順にコード化していく際に、或るノードXに直接結合するノードYをコードするとき該コード化位置以前に既にノードYをコード化していた場合には、ノードYのノードコードの代わりに閉環コードを使用し、該コード化位置以降にノードYに直接結合するノードとしてノードXが現れたときに該ノードXのノードコードの代わりに前記ノードYに使用した閉環コードと同じ閉環コードを使用し、且つ、閉環コードで表現されたノードについてはそれを起点とする以降のノードのコード化を省略するようにしたことを特徴とする。

【0019】また、本発明の有機化合物の部分構造のコード化方法は、前記第3ステップにおいて、1つの部分構造のコード化において複数の閉環コードを使用する必要がある場合、予め定められたそれぞれ異なる複数の閉環コードから循環的に1つずつ閉環コードを選択して使用することを特徴とする。

【0020】また、本発明の有機化合物の部分構造のコード化方法は、前記第2ステップにおいて、入力ファイルにケミカルシフト値が記述されている元素をフォーカスノードとし、且つ、前記第6ステップにおいて、生成した部分構造のコードと共にそのケミカルシフト値を出力ファイルに出力することを特徴とする。

【0021】また、本発明の部分構造コード化プログラムを記録した媒体は、コンピュータに下記のステップを実行させるための部分構造コード化プログラムが記録されていることを特徴とする。

第1ステップ；有機化合物を構成する元素の種類および各元素間の結合状態を記述した入力ファイルを読み込む。

第2ステップ；有機化合物を構成する元素のうちの1つの元素をフォーカスノードに定める。

第3ステップ；少なくとも炭素のノードに対してはその結合水素数に応じて異なるノードコードを割り当てるようにしつつ、フォーカスノードのノードコードを記述し

たフォーカス情報をルートとし、第1～第Nスフィアに含まれるノードのノードコードおよびボンドコードを記述したN個のスフィア情報をそれぞれ第1～第N階層とする木構造を作成する。

第4ステップ；木構造中のスフィア情報における各バンチ内のノードコードとボンドコードとの組の並びが所定の優先順位に沿うように木構造を正規化する。

第5ステップ；正規化された木構造をフォーカス情報、第1スフィア情報, ..., 第Nスフィア情報の順に走査して、フォーカス情報と第1スフィア情報との間に第1の区切り記号コードを置き、各スフィア情報間に第2の区切り記号コードを置き、各スフィア内のバンチ情報間に第3の区切り記号コードを置いたコードを生成する。第6ステップ；生成したコードを出力ファイルに出力する。

【0022】また、部分構造コード化プログラムがコンピュータに実行させる前記第3ステップが、炭素に加え、窒素および酸素のノードに対してもその結合水素数に応じて異なるノードコードを割り当て、それ以外の元素のノードに対しては1種類のノードコードを割り当てる手順を含むことを特徴とする。

【0023】また、部分構造コード化プログラムがコンピュータに実行させる前記第3ステップが、フォーカス情報、第1スフィア情報, ..., 第Nスフィア情報の順に木構造を作成していく際に、或るノードXに直接結合するノードYをコードするとき既にノードYをコード化していた場合には、ノードYのノードコードの代わりに閉環コードを使用し、以降にノードYに直接結合するノードとしてノードXが現れたときに該ノードXのノードコードの代わりに前記ノードYに使用した閉環コードと同じ閉環コードを使用し、且つ、閉環コードで表現されたノードについてはそれを起点とする以降のノードのコード化を省略する手順を含むことを特徴とする。

【0024】また、部分構造コード化プログラムがコンピュータに実行させる前記第3ステップが、1つの部分構造のコード化において複数の閉環コードを使用する必要がある場合、予め定められたそれぞれ異なる複数の閉環コードから循環的に1つずつ閉環コードを選択して使用する手順を含むことを特徴とする。

【0025】また、部分構造コード化プログラムがコンピュータに実行させる前記第2ステップが、入力ファイルにケミカルシフト値が記述されている元素をフォーカスノードとする手順を含み、部分構造コード化プログラムがコンピュータに実行させる前記第6ステップが、生成した部分構造のコードと共にそのケミカルシフト値を出力ファイルに出力する手順を含むことを特徴とする。

【0026】本発明にあつては、炭素(C)の場合、 CH_3 のように3つの水素と結合しているときは例えば4にコード化し、 CH_2 のように2つの水素と結合しているときは例えば3にコード化し、 CH のように1つの

水素と結合しているときは例えば2にコード化し、水素と全く結合していないときは例えば1にコード化する。また、窒素 (N) の場合、 NH_3 のように3つの水素と結合しているときは例えばJにコード化し、 NH_2 のように2つの水素と結合しているときは例えばGにコード化し、 NH のように1つの水素と結合しているときは例えばEにコード化し、水素と全く結合していないときは例えばNにコード化する。さらに、酸素 (O) の場合、 OH のように1つの水素と結合しているときは例えばLにコード化し、水素と全く結合していないときは例えばOにコード化する。このようにすると、例えば図55

(a) に示した有機化合物の全体構造において、番号1の CH_3 をフォーカスとした第4スフィアまでの部分構造は、

4 ; 3 (2 / 1 \$ 3 / O, 2) … (1)

のようにコード化される (このコード化の方法は「発明の実施の形態」の項の具体例1で説明している)。

【0027】このコードを見れば明らかなように、本発明では、フォーカスの元素をはじめとして各スフィアの元素の結合水素数とそのコードから直接判明する。即ち、コード4を見れば、フォーカスの元素が炭素 (C) であり、その結合水素数が3であることが直ちにわかる。第1スフィアの3、第2スフィアの2、第3スフィアの1、3、第4スフィアのO、2についても同様である。このように直接そのコードにより結合水素数が判明するため、最後のスフィア、つまり上記(1)のコードでは第4スフィアにおいても、結合水素数の情報が失われない。従って、図55(c)に示す有機化合物の全体構造において、番号1の CH_3 をフォーカスとした第4

スフィアまでの部分構造は、

4 ; 3 (2 / 1 \$ 3 / O, 1)

となり、上記(1)のコードと異なるものとなる。これは、HOSE Code と同じスフィア数の表記であっても、水素の結合する元素の環境をより精度よく表現できていることを示す。

【0028】なお、上述の説明では、炭素に対して1、2、3、4のコードを、窒素に対してN、E、G、Jのコードを、酸素に対してO、Lのコードを割り当てたが、他の元素のコードを含め、ユニークなコードであれば、例示したコードに限定されないことは勿論のことである。また、スフィア数の最大値Nは任意の値とすることができる。

【0029】

【発明の実施の形態】次に本発明の実施の形態の例について図面を参照して詳細に説明する。

【0030】先ず、主要な用語の意味等について説明する。

【0031】ノード；有機化合物は多数の元素から構成されている。この有機化合物を構成する個々の元素をノードと呼ぶ。但し、水素はノードに含めない。また、酸

素もそれが炭素、窒素等の他の元素と二重結合されている場合には、それ単独でノードとせず、結合先の元素と共に1つのノードを構成するものとする。従って、例えば図1(a)に例示した構造式の有機化合物においては、同図の符号1～9を付した部分が各々1つのノードとなる。

【0032】ボンド；ノードとノードとの間の結合をボンドと呼ぶ。ボンドには、単結合、二重結合、三重結合、芳香族結合等の種類がある。図1(a)の有機化合物において、例えばノード1とノード2間のボンドは単結合、ノード5とノード6間のボンドは二重結合である。

【0033】フォーカスノード；有機化合物の全体構造から或る部分構造を考える場合にその中心として注目するノードをフォーカスノードと呼ぶ。一般に水素の結合しているノードまたは炭素のノードをフォーカスノードとするが、目的に応じて他の元素をフォーカスノードとすることができる。

【0034】第1スフィア～第Nスフィアおよびバンチ；有機化合物の全体構造において或るフォーカスノードを定めた場合に、フォーカスノードに直接結合しているノードの集合を、フォーカスノードのバンチまたは第1スフィアと呼ぶ。第1スフィアの或るノードaに直接結合しているノードのうちフォーカスノード以外のノードの集合を、そのノードaのバンチと呼ぶ。第1スフィアの全てのノードのバンチの集合を第2スフィアと呼ぶ。同様に、第i-1 (i=3～N) スフィアの或るノードbに直接結合しているノードのうちそのノードbが直接結合している第i-2スフィアのノード以外のノードの集合を、そのノードbのバンチと呼ぶ。第i-1スフィアの全てのノードのバンチの集合を第iスフィアと呼ぶ。例えば図1(a)において、ノード1をフォーカスノードとした場合、各バンチおよび各スフィアは同図(b)に示すように定められ、ノード3をフォーカスノードとした場合、各バンチおよび各スフィアは同図(c)に示すように定められる。

【0035】ノードコード；ノードを構成する元素の元素記号を所定の規則に従ってコード化したものをノードコードと呼ぶ。ノードコードの変換規則表の例を図2に示す。

【0036】図2に示すように、炭素 (C) の元素記号は、基本的に結合水素数に応じたコードを割り当てる。つまり、Cのノードに対しては1を、 CH のノードに対しては2を、 CH_2 のノードに対しては3を、 CH_3 のノードに対しては4を、それぞれ割り当てる。この場合、ノードコード-1=結合水素数となる。一般に炭素は有機化合物の代表的な構成元素で頻繁に現れるため、ノードコードから直観的に結合水素数が分かるようなコードとしている。

【0037】また、図2に示すように、窒素 (N) 、酸

素(O)の元素記号も、結合水素数に応じたコードを割り当てる。つまり、Nのノードに対してはNを、NHのノードに対してはEを、NH₂のノードに対してはGを、NH₃(チャージ)のノードに対してはJを、それぞれ割り当て、Oに対してはOを、OHに対してはLを、それぞれ割り当てる。但し、これは¹H-NMRスペクトル等を対象とする場合であって、¹³C-NMRスペクトルを対象とする場合にはそれぞれ区別せずにN、Oで代表して良い。なぜなら、¹³C-NMRスペクトルのケミカルシフトにはN、Oなどへの結合水素数があまり影響を与えないため、水素数が問題にならないからである。なお、ノードの説明の箇所ですべてのように、酸素はそれが炭素、窒素等の他の元素と二重結合されている場合には、それ単独でノードとせず、結合先の元素と共に1つのノードを構成するが、その際、第1スフィアでは=Oとし、第2スフィア以降では\$とする。そして、\$の場合、結合先の元素のノードコードの直後に表記する。例えば、C=Oの場合、1\$と表記する。

【0038】その他の元素記号については、図2に示すように、Si, Cl, Brに対してはQ, X, Yを割り当て、I, Yに対しては”I”, ”Y”を割り当て、それ以外の1文字の元素記号に対しては元素記号と同じ記号のコードを割り当てる。また、2文字の元素記号については、元素記号を大文字化し前後に”を付加したものをノードコードとする。その他、水素は一般にコード化しないが、必要なときは”H”と表記し、重水素はDで表す。

【0039】ボンドコード；或るノードaのバンチの中にノードbがあるとき、ノードbとノードaとのボンドの種類をコード化したものを、ノードbのボンドコードと呼ぶ。前述したようにボンドの種類には単結合、二重結合、三重結合、芳香族結合等があり、図3のボンドコード表に示すように、例えば二重結合のボンドに対して「=」、三重結合のボンドに対して「%」、芳香族結合のボンドに対して「*」、その他の結合のボンドに対して「>」の各ボンドコードを割り当てる。単結合に対するボンドコードは無い。コード化に際して、ボンドコードは、ノードコードの前に付与される。

【0040】フォーカス情報、バンチ情報、スフィア情報；フォーカスノードのノードコードを記述した部分をフォーカス情報と呼ぶ。バンチに含まれる各ノードのボンドコードとノードコードとの組を記述した部分をバンチ情報と呼ぶ。スフィアに含まれる各バンチのバンチ情報を記述した部分をスフィア情報と呼ぶ。

【0041】閉環コード；環状化合物等の場合に、同じノードを2回以上コード化しないため、既にコード化したノードについては特殊なコードで代用し、そのノードに関してそれ以降のバンチをコード化しない。このとき使用する特殊なコードを閉環コードと呼ぶ。閉環コードには、本実施の形態では図4の閉環コード表に示すよう

な7種類がある。環が現れる毎に、&から順に、!, @, #, ?, _, ~ (チルド) という順番で使用していき、チルドまで使用すると、また&に戻って繰り返し使用する。即ち、循環的に使用する。一つの環では同じ閉環コードを2つペアとして使用して閉環結合の場所を示す。

【0042】次に、本発明にかかるコード化方法によるコードの構成例について説明する。なお、本発明にかかるコード化方法によって生成されたコードを以下HYPER Code (HYdrogen Part Emphatic Representation Code) と称す。

【0043】図5はHYPER Codeの一般的なフォーマットを示す。同図において、Fはフォーカス情報、SP1~SPNは第1スフィア情報~第Nスフィア情報である。フォーカス情報Fと第1スフィア情報SP1とは区切り記号「;」(セミコロン)で区切られ、第1スフィア情報SP1と第2スフィア情報SP2とは区切り記号「(」(左閉じ括弧)で区切られ、第2スフィア情報SP1から最終の第Nスフィア情報SPNまでの各スフィア情報間は「/」(スラッシュ)で区切られる。そして、最終の第Nスフィア情報SPNの後の区切り記号として「)」(右閉じ括弧)が使用される。

【0044】フォーカス情報Fには、フォーカスノードを構成する元素の元素記号を図2に示した変換規則に従ってコード化したノードコードNCが設定される。

【0045】第1スフィア情報SP1には、フォーカスノードのバンチ(第1スフィア)を構成する各ノードのボンドコードBCとノードコードNCの組が設定される。ボンドコードBCはノードコードNCの直前に記述されるが、単結合では前述したように省略される。フォーカスノードのバンチ(第1スフィア)を構成するノードが複数存在する場合、結合次数の高さ、ノード種、分岐度等に基づいて予め定められた優先順位に従ってボンドコードBCとノードコードNCの組が並べられる。この場合、各組間を区切る記号はなく、連続して列挙される。

【0046】第2スフィア情報SP2から最後の第Nスフィア情報SPNまでには次のような情報が設定される。即ち、第i(i=2~N)スフィア情報SPiには、第i-1スフィアを構成する各ノードのバンチ情報が、第i-1スフィアを構成する各ノードの並びと同じ順序で所定の区切り記号「,」(コンマ)を置いて並べられる。各バンチ情報には、そのバンチを構成する各ノードのボンドコードBCとノードコードNCの組が設定される。1つのバンチを構成するノードが複数存在する場合には、前述と同様な優先順位に従ってボンドコードBCとノードコードNCの組が並べられる。但し、第i-1スフィアを構成するノードであっても、そのバンチが存在しないノードに対するバンチ情報は区切り記

号「,」だけが置かれ、第*i*-1スフィアを構成するノードであってそれが閉環コードで置き換えられた場合は、上記の区切り記号「,」も省略される。

【0047】図6の優先順位表は1つのパンチ情報内でボンドコードBCとノードコードNCとの組を複数列挙する場合に適用する優先順位の例を示す。複数のボンドコードとノードコードとの組の優先順位の判定は、まずボンドコード間の優先順位に基づき決定し、ボンドコードが同じ場合にはノードコード間の優先順位に基づき決定する。

【0048】以下、幾つかの具体例を挙げてHYPER Codeの構成をより詳細に説明する。

【0049】具体例1；図1(a)に示した構造式の有機化合物において、ノード1をフォーカスノードとし、第4スフィアまでの部分構造をHYPER Codeで表記する。このとき各パンチおよびスフィアは図1

(b)に示すものとなる。フォーカスノードはCH₃であり、図2を参照するとそのノードコードは「4」であるため、フォーカス情報Fは「4」になる。フォーカスノード1のパンチはCH₂の1つのノード2からなり、第1スフィアはこのパンチだけから構成される。図2を参照すると、CH₂のノードコードは「3」であり、フォーカスノード1とのボンドは単結合である。従って第1スフィア情報SP1は「3」になる。

【0050】第2スフィアは、ノード2のパンチだけから構成され、そのパンチはCHの1つのノード3から構成される。図2を参照するとそのノードコードは「2」であり、ノード2とのボンドは単結合である。従って第2スフィア情報SP2は「2」になる。

【0051】第3スフィアは、ノード3のパンチだけから構成され、そのパンチはC=OとCH₂との2つのノード4、7からなる。図2を参照するとそのノードコードはそれぞれ「3」、「1\$」であり、ノード3とのボンドはそれぞれ単結合である。更に図6の優先順位表をみると1\$の方が3より優先度が高い。従って第3スフィア情報SP3は「1\$3」となる。

【0052】第4スフィアは、ノード4のパンチとノード7のパンチとから構成され、ノード4のパンチはCHの1つのノード5からなり、ノード7のパンチはOの1つのノード8からなる。図2を参照すると、CH、Oのノードコードはそれぞれ「2」、「O」であり、ノード4、7に対するボンドはそれぞれ単結合である。従って、第4スフィア情報SP4は「O, 2」となる。ここで、「2, O」としないのは、第3スフィア情報SP3においてノード7、ノード4の順番でノードコードが列挙されている為である。結局、全体のHYPER Codeは、

4 ; 3 (2 / 1\$3 / O, 2)
となる。

【0053】具体例2；図1(a)に示した構造式の有機

機化合物において、ノード3をフォーカスノードとし、第4スフィアまでの部分構造をHYPER Codeで表記する。このとき各パンチおよび各スフィアは図1

(c)に示すものとなる。フォーカスノードはCHであり、そのノードコードは「2」である為フォーカス情報Fは「2」になる。フォーカスノード3のパンチはCH₂, CH₂, C=Oの3つのノード2, 4, 7からなり、第1スフィアはこのパンチから構成される。図2を参照すると、その各々のノードコードは「3」,

「3」, 「1\$」であり、またフォーカスノード3とのボンドは何れも単結合である。さらに優先順位は1\$が3よりも高い。従って第1スフィア情報SP1は「1\$33」になる。

【0054】第2スフィアは、ノード2のパンチとノード4のパンチとノード7のパンチとから構成され、ノード2のパンチはCH₃の1つのノード1からなり、ノード4のパンチはCHの1つのノード5からなり、ノード7のパンチはOの1つのノード8からなる。それぞれのノード1, 5, 8のノードコードは「4」, 「2」, 「O」であり、またノード2, 4, 7とのボンドは何れも単結合である。従って、第2スフィア情報SP2は「O, 2, 4」となる。

【0055】第3スフィアは、ノード5のパンチとノード8のパンチとから構成され、ノード5のパンチはCH₂の1つのノード6からなり、ノード8のパンチはCH₂の1つのノード9からなる。それぞれのノード6, 9のノードコードは「3」, 「4」であり、ノード6のノード5に対するボンドは二重結合、ノード9のノード8に対するボンドは単結合である。従って、第3スフィア情報SP3は「4, =3, 」となる。ここで、最後の「,」は第3スフィアには第2スフィアのノード1に対応するパンチが存在しないことを示している。

【0056】第4スフィアは、一つのパンチも存在しない。従って、第4スフィア情報は「,」となる。結局、全体のHYPER Codeは、
2 ; 1\$33 (O, 2, 4 / 4, =3, /,)
となる。

【0057】具体例3；図7(a)に示す構造式の有機化合物において、ノード1をフォーカスノードとし、第4スフィアまでの部分構造をHYPER Codeで表記する。このとき各パンチおよび各スフィアは同図

(b)のようになる。ここで、第3スフィアと第4スフィアとが重複しているのは、図7(a)の有機化合物が環構造を有しており、第3スフィア中のノード5のパンチがノード6, 7になり、ノード6のパンチがノード5, 8になるからである。最終的なHYPER Codeによる部分構造表記は同図(c)のようになる。理解を容易にするために、ノード番号との対応を付記してある。第3スフィアまでは前述の具体例と同様にしてコード化できる。

【0058】第4スフィア情報では、先ず、第3スフィア中のノード5のパンチ情報をコード化する。ノード5のパンチはCHのノード6とNH₂のノード7から構成される。NH₂のノード7のノードコードは図2を参照すると、Gであり、ノード5に対するボンドは単結合である。他方、CHのノード6のノード5に対するボンドは単結合であり、そのノードコードは図2を参照すると、2である。しかし、ノード6は第3スフィア情報において既にコード化されている。そこで、閉環しているとして、図4の閉環コード表を参照し、先頭の閉環コード&をノード6のノードコード2の代わりに使用する。このとき、閉環コード&を使用中であることと、その閉環コード&をノード5とノード6との結合に使用していることを閉環コード使用記録として残しておく。結局、ノード5のパンチ情報はG&となる。

【0059】次に第3スフィア中のノード6のパンチ情報をコード化する。ノード6のパンチはCHのノード5とOHのノード8から構成される。OHのノード8のノードコードは図2を参照すると、Lであり、ノード6に対するボンドは単結合である。他方、CHのノード5のノード6に対するボンドは単結合であり、そのノードコードは図2を参照すると、2である。しかし、ノード5は第3スフィア情報において既にコード化されている。そこで、閉環しているとして、閉環コードの使用記録を参照し、ノード6とノード5との結合に使用中の閉環コードの記録があるか否かを調べる。今の場合、&が記録されているので、ノード5のノードコード2の代わりに閉環コード&を使用し、閉環コード&を使用済にする。結局、ノード6のパンチ情報はL&となる。

【0060】なお、環が複数あるときは、1つ目の環は&を2つペアにして使用し、次の環は!を2つペアにして使用するというように、7種類の閉環コードを順番に使用していき、最後の閉環コード「チルド」を使った後は、また&に戻って繰り返し使用する。例えば図8

(a)に示すような3つの環を持つ構造（これは説明の便宜上想定した仮想構造である）において、ノード2をフォーカスノードとして第6スフィアまでの部分構造を表記する場合、&, !, @の3つがその順番で使用され、同図(b)に示すような部分構造表記となる。このとき、同図(c)に示すように、閉環コード&はノード4とノード5が結合していることを、閉環コード!はノード6とノード7が結合していることを、閉環コード@はノード8とノード9とが結合していることを、それぞれ示す。なお、HOSE Codeのように1種類の閉環コード（例えば&だけ）しか使用しない場合には、図8(b)の部分構造表記は同図(d)に示すものとなり、どのノード同士が結合しているのか不明となる。

【0061】次に、HYPER Codeを自動的に生成する装置および方法の実施例について説明する。

【0062】図9はHYPER Code自動生成装置

の一実施例のブロック図である。この例のHYPER Code自動生成装置は、中央処理装置100と、これに接続されたメモリ（主記憶）101、プログラム記録媒体102、入力ファイル103、出力ファイル104、入力装置105、表示装置106およびプリンタ107とから構成されている。

【0063】入力装置105はキーボードやマウス等で構成され、HYPER Code自動生成装置に対する各種の指示やデータ等を入力するために使用される。表示装置106はCRTディスプレイ等で構成され、HYPER Code自動生成装置から利用者に出力すべきデータを表示するために、同じくプリンタ107はHYPER Code自動生成装置から利用者に出力すべきデータを印字出力するために、それぞれ使用される。

【0064】入力ファイル103は、有機化合物を構成する元素の種類および各元素間の結合状態を記述したファイルである。この入力ファイル103の一例を図10に示す。この例の入力ファイル103は結合表と呼ばれるものであり、ヘッダ200と、カウンタライン201と、アトムブロック202と、ボンドブロック203と、シフトブロック204とを含んでいる。ヘッダ200には、この結合表で示される有機化合物の名称等が設定される。今の場合、対象となる有機化合物は、ETHYL(4-METHOXYPHENYL)ACETATEである。その構造式を図11に示す。なお、図11において、1~14はアトム番号である。また、カウンタライン201には、この有機化合物を構成する元素の数（今の場合14）、元素と元素とが結合している数（今の場合14）等が設定される。

【0065】アトムブロック202とボンドブロック203とが結合表の主要部分であり、アトムブロック202では当該有機化合物を構成する元素の種類等が設定され、ボンドブロック203では各元素間の結合状態等が設定される。アトムブロック202はその各行が1つの元素に対応しており、先頭から順にアトム番号1の元素、アトム番号2の元素、…、アトム番号14の元素を記述している。各行の先頭から3つまでの数値は当該元素の座標値を示し、その次の記号「C」等が当該元素の種類を示す。以降、各種のデータが設定されるが、本発明に直接関係しないので省略する。ボンドブロック203はその各行が1組の元素間の結合を記述しており、各行の先頭から順に、第1元素のアトム番号、第2元素のアトム番号、結合状態、…となっている。結合状態は、1が単結合、2が二重結合、3が三重結合、4が芳香族結合を示す。例えば、1行目は、アトム番号2の元素とアトム番号1の元素とが二重結合していることを、2行目は、アトム番号3の元素とアトム番号1の元素とが単結合していることを、それぞれ示す。

【0066】最後のシフトブロック204は、この有機化合物に対するスペクトル測定の結果得られたスペクト

ルデータに関する情報が設定されている。このうち、「NHSF 1」が先頭についている行は、アトム番号とケミカルシフト値とを記述する。例えば、1行目は、アトム番号3の元素の第1および第2ケミカルシフト値が3, 5400であることを示す。なお、その後ろには強度(Intensity)等が付記されるが、本発明と直接関係しないので省略する。また、「NMSC 1」が先頭についている行は、測定共鳴周波数を示し、今の場合、90MHzである。「N SLV 1」が先頭についている行は、測定溶媒の種類を示し、今の場合、クロロホルム(CDCL₃)である。「N REF 1」が先頭についている行は、当該データが記載されていた文献名(出典名)を示す。MENDはシフトブロックの終了である。

【0067】再び、図9を参照すると、出力ファイル104は、HYPER Code自動生成装置で生成されたHYPER Codeを格納するファイルである。また、プログラム記録媒体102は、中央処理装置100で実行すべき部分構造コード化プログラムを記憶する媒体であり、このプログラムは実行時にメモリ101にロードされ、中央処理装置100で実行される。

【0068】図12は図9の装置の構成を機能ブロック図化したものである。同図において、図9と同一符号は同一部分を示し、300はプログラム記憶媒体102に記憶されたプログラムと中央処理装置100とメモリ101とで実現されるコード生成機構である。このコード生成機構300は、パラメータ入力手段301と、パラメータ格納エリア302と、入力ファイル読み込み手段303と、入力ファイル格納エリア304と、木構造生成手段305と、木構造生成エリア306と、木構造正規化手段307と、コード化手段308と、出力手段309と、コード格納エリア310と、図2に示した変換規則表311と、図3に示したボンドコード表312と、図4に示した閉環コード表313と、図6に示した優先順位表314とで構成されている。

【0069】図13はコード生成機構300の処理の概要を示すフローチャートである。コード生成機構300は起動されると、まず、パラメータ入力手段301によって入力装置105を通じて利用者から各種のパラメータを入力し、パラメータ格納エリア302に格納する(S1)。パラメータとしては、例えば、図12に示すように、入力ファイル名3021、出力ファイル名3022、スフィア数3023、変換モード3024、フォーカス選択モード3025がある。スフィア数3023では、これから作成しようとするHYPER Codeの最大スフィア数を定める。変換モード3024では、変換規則表311を使用したノードコードへの変換時に、窒素および酸素をそれぞれN, Oで代表させるモードにするか、結合水素数に応じてコード化するモードにするのかを定める。本実施例では窒素および酸素もその

結合水素数に応じてコード化するモードが指定されている場合を採り上げる。なお、何れのモードであっても炭素についてはその結合水素数に応じてコード化されるのは勿論のことである。フォーカス選択モード3025では、どのようなノードをフォーカスノードとするかを定める。前述したように、水素の結合しているノードや炭素のノードをフォーカスとすることができ、また目的に応じて他の元素をフォーカスとすることができるため、このフォーカス選択モード3025でフォーカスノードの指定を行う。本実施例では、入力ファイル103のシフトブロック204にケミカルシフト値が挙げられている元素、つまり図10の場合はアトム番号3, 7, 8, 9, 10, 12, 13, 14の元素のノードをフォーカスノードとするモードが指定されている場合を採り上げる。

【0070】次にコード生成機構300は、入力ファイル読み込み手段303によって、入力ファイル名3021に該当する入力ファイル103を読み込み、入力ファイル格納エリア304に格納する(S2)。

【0071】次にコード生成機構300は、フォーカスノードの数だけステップS3～S6の処理を繰り返し実行し、最後にステップS7を実行して、コード生成処理を終了する。

【0072】ステップS3においてコード生成機構300は、木構造生成手段305により、フォーカス選択モード3025に基づき入力ファイルから1つのフォーカスノードを選択する。今の例の場合、シフトブロック204にケミカルシフト値が挙げられているアトム番号3, 7, 8, 9, 10, 12, 13, 14の元素のノードを1つずつフォーカスノードとして選択していく。

【0073】ステップS4においてコード生成機構300は、木構造生成手段305により、入力ファイルのアトムブロック202およびボンドブロック203、変換規則表311、ボンドコード表312および閉環コード表313を参照して、フォーカスノードのノードコードを記述したフォーカス情報をルートとし、第1～第Nスフィアに含まれるノードのノードコードおよびボンドコードを記述したN個のスフィア情報をそれぞれ第1～第N階層とする木構造を、木構造生成エリア306に作成する。

【0074】ステップS5においてコード生成機構300は、木構造正規化手段307により、木構造生成エリア306に作成された木構造中のスフィア情報における各パンチ内のノードコードとボンドコードとの組の並びが優先順位表314の定める優先順位通りになるように木構造を正規化する。このような木構造の正規化を行う理由は、全く同一の部分構造からは一意なHYPER Codeが生成されるようにするためである。

【0075】ステップS6においてコード生成機構300は、コード化手段308により、木構造生成エリア3

10

20

30

40

50

06に作成された正規化後の木構造に基づきHYPER Codeを作成し、コード格納エリア310に格納する。

【0076】ステップS7においてコード生成機構300は、出力手段309により、コード格納エリア310に格納されたHYPER Codeを、出力ファイル名3022に該当する出力ファイル104に出力する。

【0077】図14は図13のステップS4の詳細を示すフローチャートである。このステップS4は木構造生成手段305で実行される。以下、木構造生成処理を、その理解を容易にするために、図10のシフトブロック204におけるアトム番号7の元素のノードがフォーカスノードとして選択された場合を例にして詳細に説明する。なお、本実施例では、例えば図15に示すようにデータ域400とm個のポインタ設定域401-1~401-mとから構成される、セルと呼ばれるデータ構造を、順次にリンクしていくことにより、木構造を表現する。このとき、木構造の根（ルート）に当たる部分のセルをルートセルと呼ぶ。また、或るセルAの或るポインタ設定域401-i（i=1~m）に設定されたポインタによってそのセルAに別のセルBがリンクされている場合、セルAとセルBとの関係において、セルAを親セル、セルBを子セルと呼ぶ。

【0078】まず、木構造生成手段305は、アトム番号7の元素の元素記号「C」をアトムブロック202を参照して求め、次いでボンブロック203からアトム番号7と結合する他の元素との結合状態を調べて結合水素数を決定し、変換モード3024を考慮して変換規則表311を参照し、アトム番号7の元素のノード、つまりフォーカスノードのノードコードを作成して、このノードコードを格納したルートセルを木構造生成エリア306に作成する（S11）。ボンブロック203を参照すると、アトム番号7の炭素は、アトム番号4、9の元素とそれぞれ芳香族結合しているため結合水素数は1である。従って、フォーカスノードのノードコードは「2」となる。木構造生成エリア306に作成されるルートセルでは、図15のデータ域400にフォーカスノードのノードコード「2」が設定される。なお、本実施例では、ケミカルシフト値等もHYPER Codeと共に出力するため、この時点でルートセルのデータ域400に、シフトブロック204の該当するケミカルシフト値「7.2000」と、アトム番号7と、溶媒コード「CD」と、共鳴周波数「90h」とが設定される。また、各ポインタ設定域401-1~401-mは子セルとのリンク時に、401-1から順番に使用されるが、この時点では各ポインタ設定域401-1~401-mはNULL値である。

【0079】次に木構造生成手段305は、内部変数Lに階層レベル1を設定し（S12）、同じく内部変数である集合Uに、階層レベルL-1のセルのノードのア

ム番号を格納する（S13）。ここで、階層レベルL-1が0の場合はルートセルを示す。従って、最初は集合Uにフォーカスノードであるアトム番号7が格納される。

【0080】次に木構造生成手段305は、集合Uからアトム番号7のフォーカスノードを取り出し（S14）、この取り出したフォーカスノードのパンチを構成するノードをボンブロック203およびアトムブロック202を参照して求めて、内部変数である集合Vにそのノードを特定するアトム番号を格納する（S16）。アトム番号7のフォーカスノードの場合、ボンブロック203をその先頭行より順に参照すると、6行目でアトム番号4の元素が求まり、8行目でアトム番号9の元素が求まるため、集合Vはアトム番号4、9のノードとなる。

【0081】次に木構造生成手段305は、集合Vからアトム番号4のノードを取り出し（S17）、このアトム番号4の元素の元素記号「C」をアトムブロック202を参照して求め、次いでボンブロック203からアトム番号4と結合する他の元素との結合状態を調べて結合水素数およびフォーカスノードに対するボンコードを決定し、ボンコード表312および変換規則表311を参照してアトム番号4のノードのボンコード「*」とノードコード「1」の組を作成し、これらを図15のデータ域400に格納したセルを木構造生成エリア306に作成し、親セルであるフォーカスノードのルートセルにリンクする（S19）。つまり、親セルの第1番目のポインタ設定域401-1にアトム番号4のノード用のセルへのポインタを設定する。

【0082】次に木構造生成手段305は、集合Vからアトム番号9のノードを取り出し（S17）、前述と同様にしてボンコード「*」とノードコード「2」の組を作成し、そのノード用のセルに格納して親セルとリンクする（S19）。

【0083】次に木構造生成手段305は、集合Vから次のアトム番号のノードを取り出すが、集合Vはその時点で空なので（S18でYES）、ステップS14に戻り、集合Uから次のノードを取り出す。しかし、この集合Uもその時点では空なので（S15でYES）、ステップS20において変数Lを+1して階層レベル2とし、ステップS21を経由してステップS13に戻り、階層レベル1のセルのノードであるアトム番号4、9を集合Uに格納する。そして、この集合Uが空になるまで、ステップS14~S19の処理を繰り返し、空になったら再び階層レベルを+1して3とし、上述した処理を繰り返す。このような処理を変数Lの値がパラメータ格納エリア302に格納されたスフィア数3023を超えるまで続ける（S21）。

【0084】図16は以上のような木構造生成手段305の処理によって木構造生成エリア306に生成された

木構造の概要を示している。同図において、括弧内に記載した数値はアトム番号を示し、その前に付記した文字列はボンドコードとノードコードとの組を示す。なお、この木構造では、親セルにリンクされている子セルが複数ある場合、親セルの第1番目のポインタ設定域401-1に設定されたポインタでリンクされている子セルを最も左側に記載している。

【0085】図16の木構造を見ると、第2階層のアトム番号5のパンチを構成するノードの一つが閉環コード&で表現されている。これは、本来はアトム番号10のノードコード「2」を記載すべきであるが、アトム番号10のノードは第2階層のアトム番号8のノードのパンチを構成するノードとして既にコード化されているため、閉環コードで置換したものである（この処理は図14のフローチャートではステップS19の処理の一部として行われる）。同様に、第3階層のアトム番号10のパンチを構成するノードが閉環コード&で表現されている。これも、本来はアトム番号5のノードコード「1」を記載すべきであるが、アトム番号5のノードは第1階層のアトム番号9のノードのパンチを構成するノードとして既にコード化されているため、閉環コードで置換されたものである。そして、この場合、アトム番号5とアトム番号10との結合に同じ閉環コードを使用するため、&がペアとして使用されている。

【0086】図17は図13のステップS5の処理の詳細を示すフローチャートである。この処理は木構造正規化手段307で実行される。木構造正規化手段307は、先ず、内部変数であるLを階層レベル1に設定し（S31）、同じく内部変数である集合Xに、階層レベルL-1のセルを格納する（S32）。階層レベルが0のときはルートセルを示す。従って、最初に集合Xにはルートセルが入る。次に木構造正規化手段307は、集合Xから1つのセル、つまりルートセルを取り出し（S33）、そのルートセルの子セルが、ボンドコードとノードコードに関し、優先順位表314に示される優先順序通りに並んでいるか否かを調べる（S35）。つまり、優先順位の高いボンドコードとノードコードを持つ子セルほど、ルートセルの第1ポインタ設定域401-1に近い方にリンクされているか否かを調べる。そして、優先順位通りに並んでいない場合には、優先順位通りに並ぶようにリンクを付け換えて子セルを並べ換える（S36）。なお、優先順位の同じボンドコードとノードコードの子セルが並んでいる場合は、この階層レベルだけでは優劣をつけられないので、更に階層レベルを下げて優劣を判定し、優先順位の高い孫セル、曾孫セル等を持つ子セルほど第1ポインタ設定域401-1に近い方にリンクされるように子セルを並べ換える。

【0087】集合Xからルートセルを取り出して処理すると、集合Xは空になるので（S34でYES）、ステップS37において変数Lを+1して階層レベル2と

し、ステップS38を経由してステップS32に戻り、階層レベル1のセルを集合Xに入れて、先と同様の処理を繰り返す。以上のような処理を繰り返し、変数Lを+1した時点でスフィア数を超えた場合に正規化処理を終了する。

【0088】例えば図16に示した木構造の場合、第1階層のアトム番号4のノードのセルには、アトム番号3, 8の2つのノードのセルがその順番にリンクされているが、アトム番号3のノードのボンドは単結合であるのに対してアトム番号8のボンドは芳香族結合であり、優先順位はアトム番号8のノードの方が高い。そこで、木構造正規化手段307では、アトム番号4のノードのセルに対するアトム番号3, 8のノードのセルのリンクを付け替え、図18(a)に示すように変更する。同様に図16中の第2階層のアトム番号5の子セルも優先順位通りにリンクされていないため、図18(a)に示すように変更する。

【0089】図19は図13のステップS6の詳細な処理例を示すフローチャートである。この処理はコード化手段308で実行される。先ず、コード化手段308は、正規化後の木構造のルートセルのデータ域400に設定されたアトム番号、溶媒コード、共鳴周波数、ケミカルシフト値をコード格納エリア310に出力する（S41）。続いて、ステップS42以降の処理を実行することにより、当該アトム番号のノードをフォーカスノードとするHYPER Codeを生成して、コード格納エリア310の前記出力した値の直後に出力する。以下、図18(a)の正規化後の木構造を例にしてその動作を説明する。

【0090】先ず、ルートセルのデータ域400に設定されたノードコード「2」を出力し（S42）、次いで区切り記号「;」を出力する（S43）。次に、内部変数である集合Yにルートセルの子セルであるアトム番号4, 9のノードのセルをこの順に格納する（S44）。次に集合Yから最初のセルであるアトム番号4のノードのセルを取り出し（S45）、そのセルのデータ域400中のボンドコードとノードコードの組「*1」を出力する（S47）。続いて集合Yから次のセルであるアトム番号9のセルを取り出し（S45）、そのセルのデータ域400中のボンドコードとノードコードの組「*2」を出力する（S47）。これで、集合Yは空になり（S46でYES）、区切り記号「(」を出力する（S48）。

【0091】次に内部変数であるLに階層レベル1を設定し（S49）、階層レベル1のセルであるアトム番号4, 9のノードのセルをその順に集合Yに格納する（S50）。次に集合Yから1つのセルであるアトム番号4のノードのセルを取り出し（S51）、そのセルの子セルであるアトム番号8, 3のセルをその順に集合Zに格納する（S54）。そして、この集合Zから1つのセル

であるアトム番号8のセルを取り出し(S55)、そのセルのデータ域400のボンドコードとノードコードとの組「*2」を出力する(S57)。次に集合Zから残りのセルであるアトム番号3のセルを取り出し(S55)、そのセルのデータ域400のボンドコードとノードコードとの組「3」を出力する(S57)。この時点で集合Zが空となり(S56でYES)、区切り記号「,」を出力する(S58)。

【0092】次に、集合Yから残りのセルであるアトム番号9のノードのセルを取り出し(S51)、そのセルの子セルであるアトム番号5のノードのセルを集合Zに格納する(S54)。そして、集合Zからアトム番号5のノードのセルを取り出し(S55)、そのセルのデータ域400のボンドコードとノードコードとの組「*1」を出力する(S57)。この時点で集合Zが空となり(S56)、区切り記号「,」を出力する(S58)。

【0093】ステップS51に戻って集合Yから次のセルを取り出そうとするが、この時点で集合Yは空である(S52でYES)。そこで、最後に出力した区切り記号「,」を区切り記号「/」に変更し(S59)、変数Lを+1して階層レベル2とし(S60)、ステップS61を経由してステップS50に戻り、第2階層のセルであるアトム番号8, 3, 5のノードのセルをその順に集合Yに格納し、ステップS51~S58を集合Yが空になるまで繰り返す。これにより、*2, 1\$, *&O, が出力され、その直後に最後に出力した区切り記号「,」が「/」に変更される。次に、集合Yに第3階層のアトム番号10, 1, 閉環コード&, アトム番号11のノードのセルが格納され、ステップS51~S58を集合Yが空になるまで繰り返すことにより、*&, O, 4, が出力され、その直後に最後に出力した区切り記号「,」が「/」に変更される。ここで、閉環コード&のセルが集合Yから取り出されたときはステップS53の判断結果がYESとなり、ステップS54~S58がスキップされるため、区切り記号「,」も出力されない。次に、集合Yに第4階層のセル(&, アトム番号6, 13に対応するセル)が格納されて処理されることにより、3, が生成され、その後、ステップS59で最後の「,」が「/」に変更される。ここで、3の後の「,」はアトム番号13のノードに子セルが存在しないため、集合Zが始めから空となり、直ちにステップS58に進んだことによる。次に、第5階層のアトム番号12のセルが集合Yに格納されて処理されることにより、4/が生成される。そして、この処理の後に変数Lを+1するとパラメータ格納エリアに格納されたスフィア数3023と等しくなり、ステップS62で最後に出力された区切り記号「/」が区切り記号「)」に変更される。

【0094】なお、有機化合物全体のノード数、フォーカスノードの位置および利用者から指定されたスフィア

数3023の値によっては、木構造生成手段305でセルが生成される階層レベルの最大値がスフィア数3023に達しない場合があるため、ステップS63で階層レベルLのセルが存在するか否かを調べ、存在しない場合には、ステップS64において区切り記号「/」を(スフィア数-L)+1個出力してステップS62に進むようにしている。

【0095】以上のようなコード化手段308による処理によって、図18(a)の木構造から同図(b)に示すHYPER Codeが生成される。

【0096】図12の出力手段309は、コード格納エリア310に生成されたHYPER Codeを、出力ファイル104に出力する。このときの出力ファイルの内容例を図20に示す。同図において、501, 502, 503, 504, 505, 506, 507, 508がそれぞれアトム番号3, 7, 8, 9, 10, 12, 13, 14のノードをフォーカスノードとするHYPER Codeとその付加情報(アトム番号, 溶媒コード, 共鳴周波数, ケミカルシフト値)とを示している。なお、生成したHYPER Codeおよび付加情報は出力ファイル104に出力する以外に、表示装置106またはプリンタ107或いはその両方に出力するようにしても良い。

【0097】次に、本発明の他の実施例について図面を参照して詳細に説明する。

【0098】図21は、入力ファイルからHYPER Codeを得るまでの本発明の実施例における処理の流れを示す。同図に示すように、本実施例においては、I. 初期ツリー作成処理、II. ツリーの正規化処理、III. 閉環シンボルの並べ替え処理、IV. 閉環シンボル(&)を各閉環シンボル(&を含めて7種類)に変更する処理、V. ツリーからHYPER Codeを生成する処理の順で、処理が進められる。以下、各場合毎に説明する。なお、本実施例の説明においては、ボンドコードとノードコードとの組をシンボルと呼んでいる。また、閉環シンボルは閉環コードと同じ意味で使っている。

【0099】I. 初期ツリー作成処理

この初期ツリー作成処理では、入力ファイル中の或るノードをフォーカスノードとし、図12の木構造生成手段305において説明したような木構造を作成する。即ち、図22(a)に示すような構造の有機化合物(ナフタレン誘導体)において、原子番号7のノードをフォーカスノードとする場合、同図(b)に示すような初期ツリーを作成する。なお、同図(b)のツリーの各ノードには、シンボル以外に丸数字によってノードIDが付与されている。本実施例の場合、このようなツリーは、コンピュータ内では、図23に示すような表(ノードテーブル構造体)として保存される。

【0100】II. ツリーの正規化処理

この処理については以下の項目に分けて説明する。

II-1. ツリーの正規化ルール

1. シンボルの記述順序

2. 環式化合物の正規化ルール

II-2. 正規化アルゴリズムの考え方

II-3. ツリーの正規化処理フローチャート

【0101】II-1. ツリーの正規化ルール

1. シンボルの記述順序

① 分岐原子はそのシンボルの優先順位に従うように、優先順位のより高い分岐シンボルから記述する。シンボルの優先順位は、図6において説明したように、まずボンドコード間の優先順位に基づき決定され、ボンドコードが同じ場合にはノードコード間の優先順位に基づき決定される。つまり、図24(a)の左側に示すように、フォーカスノードであるCHから見て分岐原子となる2つのノードが同一原子CHであるならば、同図の右側に示すようにボンドコードの順位で、優先順位が決まる。

② 同一シンボルで優先順位が同じ場合は、次の下位スフィアを見て、下位スフィアでシンボルが優先順位に従って並ぶように記述する。例えば、図24(b),

(c)の左側に示すような構造の場合は、それぞれ同図の右側に示すように記述する。

③ 分岐がない場合は分岐がある方を優先する。即ち、図24(d)の左側に示すような構造の場合、右側に示すように記述する。

④ 現に記述しようとするスフィアのシンボルの優先順位および分岐状態が同じなら、更に下位のスフィアに対して上記のルールを適用して、最上位の分岐が先に記述されるようにする。例えば、図24(e)の左側に示す構造の場合、第1スフィアの順位は第2スフィアで決められないので、更に下位スフィア(第3スフィア)を見て、同図の右側に示すように記述する。

【0102】2. 環式化合物の正規化ルール

例えば図25に示すような偶数環の存在する多環式化合物において、●のノードをフォーカスとした場合に、 α , β のどちらから先に環構造を回るかによってツリーのノード分岐状態およびノードのシンボルが変わることがある。即ち、図26(a)の構造と、その構造中の2cと2bとを交換した図27(a)の構造とでは、図26(b), 図27(b)に示すように、網かけを施した部分が相違する。なお、図26(b), 図27(b)において、ノード間を結ぶ破線は閉環原子対を示す。

【0103】そこで、本実施例においては、環式化合物の正規化ルールとして次のようなルールを導入する。

【0104】今、図28に示すように、ノード α , β 配下に偶数閉環があり、 α , β が同一シンボルで、シンボルを見ただけではどちらを先に記述するか分からない場合、まず、図29に示すように、 $\alpha-\beta$ 順位と $\beta-\alpha$ 順位の2つの最終スフィアまでのツリーを生成する。そして、図29に示すように、偶数閉環スフィア以降で、上

記2つのツリーの高い側のノードを、比較①, 比較②, 比較③というように上位スフィアから順に比較し、最初に優劣が決した結果に応じて、上記2つのツリーの一方を選択する。図29の場合、比較②で、 $\beta-\alpha$ 順位が優位となり、 $\beta-\alpha$ 順位のツリーを選択する。

【0105】上記ルールを図26および図27の2つのツリーに適用すれば、図27のツリーが採用される。

【0106】II-2. 正規化アルゴリズムの考え方

○正規化の実現上の問題点

対称性が高い化合物ほど、各ノードが同順位になり、正規化が難しくなっていく。有機化合物は特にその構成元素の種類が少なく(主にC, H, O, N, S)、各ノードは同順位になりやすい。対称性の高い構造例を図30に示す。本実施例では、このような問題をクリアするために、以下の方法により正規化処理を実現する。

【0107】○正規化アルゴリズムの考え方

次の2つの基本事項を見いだせる。

①どのような複雑なツリー構造でも、図31(a)に示すような最小単位から構成される。

②上位スフィアの同順位ノードの並びはその下位スフィアノード(の優先順位)により決定される(上位スフィアノードの並びは下位スフィアノードの優先順位に影響されうる)

【0108】以上の基本事項を踏まえ、本実施例では、図31(b)に示すように、ツリーの最下位スフィアの左端ノードから右端ノードへ、そして上位スフィアノードに対して、再帰的にノード順位決定処理を適用するという正規化アルゴリズムを採用する。なお、図31

(b)に示す①~⑤の数字は、図31(a)の親ノードXの遷移する順番を示す。最後にフォーカスノードがXになる。具体例を以下に示す。

【0109】例えば、図32(a)に示す化合物構造の変換を考えてみる。初期ツリーが図32(b)のようになったとして、複数の子ノードを持つノードが親ノードとなるようにすれば、図32(b)の丸数字を付けたノードが順番に親ノードXになる(親ノードになる順番は①→②→③→④→⑤)。

【0110】親ノード①の場合、子ノードの2つは同一順位で、しかも子ノードから先の分岐もないので、入れ替えは必要ない(図32(c)参照)。

【0111】親ノード②の場合、子ノード3, 4, 2を2, 3, 4に入れ替える(図32(d)参照)。

【0112】親ノード③の場合、子ノード4, 3を3, 4に入れ替える(図32(e)参照)。

【0113】親ノード④の場合、子ノード4, 3, 2を2, 3, 4に入れ替える(図32(f)参照)。

【0114】親ノードが⑤にきたとき、ツリー状態は図33(a)のようになっている。

【0115】ここで、図33(b)に示すように、親ノード⑤の子ノード同士を比較すると(比較1)、同一シ

ンボルで同一順位なので、その配下の子ノードを比較する（比較2）。

【0116】比較2では、図33（c）に示すように、左端のノードから順に順位が決まるまで、比較2-1→比較2-2→比較2-3というようにノードを比較していく。

【0117】この例の構造の場合、比較2で順位が決まらないので、図34（a）に示すように、更に下位スフィアを調べる（比較3）。

【0118】この比較3における最初の比較3-1で、図34（a）に示すB側の方が優位となる。よって、A、Bノードを入れ替えて、図34（b）のツリーを得る。

【0119】ノード⑤はフォーカスノードであるので、図34（b）のツリーが正規化ツリーとなる。

【0120】HYPER Codeは、図34（b）の正規化ツリーに対し、上から下へ、左から右へ、文字列になおしていくことで生成できる。図34（c）はこのとき生成されるHYPER Codeを示す。

【0121】このように、それぞれの分岐ごとで、下位スフィアノードの並べ換えが終わった状態になっていれば、新たな上位スフィアに対する順位決定は、そのスフィアのA、B配下の左端から順に比較していけば良いことになるので、複雑なノード順位決定処理が、かなり簡潔なものとなる。

【0122】II-3. ツリーの正規化処理フローチャート
上述した正規化アルゴリズムに基づく、ツリーの正規化処理のフローチャート例を図35～図39に示す。

【0123】図35は全体の概略フローチャートであり、ステップS101～S103から構成される。ステップS101では、親ノードスフィア変数に0をセットする。ステップS102では、フォーカスノードを親ノードXとする。そして、ステップS103において、親ノードX配下の並べ替え処理を行う。

【0124】図36は親ノードX配下の並べ替え処理の詳細なフローチャートであり、ステップS201～S209から構成される。この親ノードX配下の並べ替え処理は、図35のステップS103で呼び出される他、自分自身のステップS202からも呼び出される（再帰呼び出し）。ステップS201～ステップS203は、初期値 $k=1$ 、ループ継続条件 $k \leq n$ 、ループ増分 $k \leftarrow k+1$ の条件の下で、親ノードX配下のノード並べ替え処理を繰り返し実行する部分である。ステップS204、S209のループiは、初期値 $i=1$ 、ループ継続条件 $i < n$ 、ループ増分 $i++$ の条件の下で、ステップS205～S208のループjを繰り返し実行する部分であり、ループjは、初期値 $j=i+1$ 、ループ継続条件 $j \leq n$ 、ループ増分 $j++$ の条件の下で、DTに親ノードXと子ノードi、jをセットする処理（ステップS206）と2ノードi（A）、j（B）配下の入れ替え処理

（ステップS207）とを繰り返し実行する部分である。ここで、DTとは、図36に付記するような構造を持つデュアルノードテーブルを指し、図33において説明した親ノードの子ノード同士の比較結果を保存しておくテーブルである。以上のような処理によって、ツリーの最下位スフィアの左端ノードから右端ノードへ、そして上位スフィアに対して、再帰的にノード順位決定処理が適用され、正規化ツリーが生成される。

【0125】参考までに、図40に、図36のフローチャートで示される親ノードX配下のノード並べ替え処理のアルゴリズムスケルトンを示してある。

【0126】次に図37は、2ノードA、B配下の入れ替え処理の詳細なフローチャートであり、ステップS301～S309から構成される。この2ノードA、B配下の入れ替え処理は、図36のステップS207から呼び出される他、自分自身のステップS307からも呼び出される（再帰呼び出し）。ステップS301ではDT内のA、Bのどちらが優先順位が高いかを決め、DTの順位フラグに順位を設定する。ステップS302では、ステップS301の結果を判定し、DTの順位が決定できたらステップS303へ進み、そうでなければステップS304へ進む。ステップS303では、決定されたDT順位に従ってツリーを並べ替え、ステップS304へ進む。ステップS304では、「親ノードスフィア<最終スフィア、かつ、DTの順位が未定」という条件が成立しているか否かを判定し、成立していればステップS305へ進み、そうでなければステップS308へ進む。ステップS305では、DTに次スフィアノードを追加し、次いでステップS306で、親ノードスフィアに1を足し、次いでステップS307で、自分自身を呼び出す。他方、ステップS308に進んだ場合は、順位決定スフィアに偶数閉環があるかどうかを調べ、なければ今回の処理を終了し、あれば、ステップS309に進む。このステップS309では、別ツリーを作成して優先度の高いツリーを取得する。

【0127】参考までに、図41に、図37のフローチャートで示される2ノードA、B配下の入れ替え処理のアルゴリズムスケルトンを示してある。

【0128】次に図38は、DT順位に従ってツリーを並べ替える処理の詳細なフローチャートであり、ステップS401～S408から構成される。このDT順位に従ってツリーを並べ替える処理は、図37のステップS303から呼び出される他、自分自身のステップS407からも呼び出される（再帰呼び出し）。ステップS401では、DT順位で右側（A側）優先か否かを判定し、右側優先であればステップS402へ進み、そうでなければ今回の処理を終了する。ステップS402では、ツリー上のA、B配下を入れ替える。ステップS403では、ツリー全体の閉環ノードを入れ替える。ステップS404では、DTの最上位ノードの親ノードをX

とする。そして、ループkを繰り返す。ここで、ループkは、初期値 $k=1$ 、ループ継続条件 $i \leq n$ 、ループ増分 $k++$ の条件の下で、親ノードスフィアに1を足すステップS406、ノードk配下の並べ替え処理を行うステップS406を、繰り返すループである。

【0129】参考までに、図42(a)に、図38のフローチャートで示されるDT順位に従ってツリーを並べ替える処理のアルゴリズムスケルトンを示してある。また、図42(b)はこの処理によってDT構造体に入るノードのイメージを示す。

【0130】次に図39は、別ツリーを作成して優先度の高いツリーを取得する処理の詳細なフローチャートであり、ステップS501～S511から構成される。この別ツリーを作成して優先度の高いツリーを取得する処理は、図37のステップS309から呼び出される。ステップS501では、従来の(元の)ツリー(TT1)をコピーしてTT2とする。ステップS502では、従来の(元の)デュアルノードテーブル(DT1)をコピーしてDT2とする。ステップS503では、DT2順位を反転させる。ステップS504では、DT2順位に従ってTT2を並べ替える。ステップS505では、TT1とTT2に違いがあるかどうかを調べ、違いがなければ処理を終了し、あればステップS506以降へ進む。ステップS506では、TT1の最終スフィアまでをDT1に追加する。ステップS507では、TT2の最終スフィアまでをDT2に追加する。ステップS508では、DT1、DT2の各左側より、DD(ダブルDT)を作成する。ステップS509では、DD内の左右の順位を決定する。ステップS510では、DDの順位が右側優先かどうかを判定し、右側優先でなければ処理を終了し、右側優先であればステップS511へ進む。ステップS511では、TT2をTT1に、DT2をDT1とする。

【0131】参考までに、図43に、図39のフローチャートで示される別ツリーを作成して優先度の高いツリーを取得する処理のアルゴリズムスケルトンを示してある。また、図44は、図39のステップS508で行われる、DT1、DT2の各左側より、DDを作成する処理のイメージを示し、図45は、図39のステップS509で行われるDD内の左右の順位決定処理のイメージを示す。

【0132】III. 閉環シンボルの並べ替え処理
HYPER Codeを利用して、二つの化学構造間の共通な部分構造を見つけ出したり、部分構造を組み立てていく場合、化学構造とそれに対応するコードは1対1になっていなければならない。これは、環構造を含む有機化合物においても必要なことであり、同じ構造の有機化合物からは同じ構造の正規化ツリーを生成すること必要がある。しかし、環構造を含む有機化合物の場合、同じ構造から2つの異なる正規化ツリーが生成される可能

性がある。以下、この点について説明する。

【0133】例えば、図46(a)に示す化合物(ナフタレン誘導体)の原子番号7をフォーカスにした場合のHYPER Codeを導出することを考える。原子番号7を起点(フォーカス)にしたときの、HYPER Codeのシンボルの初期ツリーは、図46(b)に示すようになる。次に、この初期ツリーの正規化を行うことにより、図46(c)に示すような正規化ツリーが得られる。次に、図47(a)に示すように、上記の正規化ツリー図46(c)中の、対応している&(閉環シンボル)同士を、それぞれ閉環を表す別な記号に置き換える。HYPER Codeは、図47(a)のツリーを上から下へ、左から右へ向かって、コード化していくことにより、図47(b)のように生成される(6スフィアのHYPER Code)。

【0134】さて、ここで、一つの問題がある。図46(c)に示した正規化ツリーにおいて、原子番号10と原子番号9配下のツリー状態は同じなので(原子番号10→8→5と9→6→3のシンボルの並びが同じ)、図48(a)、(b)に示すような2つの正規化ツリーが生成される可能性があることである。そして、この図48(a)、(b)の正規化ツリーより、対応している閉環シンボル同士を、それぞれ別の閉環シンボルに替えると、図48(c)、(d)に示すようになり、それらから生成されるHYPER Codeは、図48(e)、(f)に示すものとなる。即ち、異なったHYPER Codeが得られてしまう。

【0135】そこで本実施例においては、有機化合物の構造とHYPER Codeとの間の一意性をより確実に確保するために、閉環シンボルノードに着目して、正規化ツリーのノードを再度並べ替える。以下、この点について詳細に説明する。なお、以後の説明を容易にするために、'閉環シンボル対'という用語を用いる。例えば、図49(a)に示す構造(ベンゼン)のHYPER Code変換を例にとれば、閉環シンボル対は図49(b)に示す、2つの*&シンボルのノードである。また、図49(c)に示すように環が2つある構造の場合、図49(d)に示すように閉環シンボル対は2組(計4つの&)となる。

【0136】さて、図46(a)に示したナフタレン誘導体の例に戻る。

【0137】ツリーの正規化処理によって、前述した図48(a)、(b)のうち、(a)側のツリーが生成されたものとする。このとき、図50(a)に示すように、閉環シンボル対(対になっている&同士)を結ぶ線を描いてみる。この線を、'閉環シンボル連結線'と呼ぶことにする。本実施例では、この閉環シンボル連結線が交差しており、且つ、2つの閉環シンボルが交換可能ならば、これらの閉環シンボル連結線が平行になる(交差しないようになる)ようにツリーのノードを入れ替え

ることで、有機化合物の構造とHYPER Codeとの間の一意性を確保する。

【0138】図50(a)の場合、2つの閉環シンボル連結線(閉環シンボル対を結んだ線)が交差しており、かつ、2つの閉環シンボル(ノードID11とノードID12)が交換可能である。従って、これらの閉環シンボル連結線が平行になるようにツリーのノードを入れ替える。この例ではノードID4とノードID3の並びを逆(入れ替える)にする。この結果、図50(a)のツリーは図50(b)のようになる。

【0139】ここで、本実施例では、ツリー構造は前述した図23に示すようなノードテーブル構造体としてコンピュータ内に格納されている。従って、上記の閉環シンボル連結線の交差検出は、このノードテーブル構造体より、各ノードのシンボル、ノード間の親子関係等を参照して行われる。因みに、図23に示すノードテーブル構造体は図46(a)の化学構造と同じ図22(a)の化学構造、つまりナフタレン誘導体のものである。

【0140】また、2つの閉環ノードが交換可能か否かは以下のようにして検出することができる。今、交換可能性を検出しようとする2つの閉環ノードを図51

(a)に示す&1、&2とし、&1、&2への最初の分岐ノードを α 、 β とする。まず、&1、&1への最初の分岐ノード α 、 β に対して、ノード順位の決定処理(II.1.ツリーの正規化 参照)を行う。&1、&2のスフィアまで調べてみて、 α と β の順位が決定できない(同一順位)ならば、 α 配下と β 配下のツリー構造は同一ということになる。つまり、 α と β は交換可能である。従って、&1、&2は、 α と β を交換することにより、交換可能である。なお、その後、&1と&2に対してもう一方の閉環シンボルを調べ、図51(b)に示すように閉環シンボル連結線が交差しているならば、図51(c)に示すように、 α と β を交換することになる。

【0141】以上のような手法を使うことにより、図50(a)に示したツリーから図50(b)に示すツリーが得られる。そして、この図50(b)のツリーより、一方の&対を別な閉環シンボル!に置き換えると、図52(a)に示すツリーとなる。このツリーより、HYPER Code(6スフィア)を生成すると、図52(b)のようになる。

【0142】V. ツリーからHYPER Codeを生成する処理

ツリーからHYPER Codeへの変換は、ツリーのシンボルをスフィアの左から右へ、上位スフィアから下位スフィアに向かって文字列に並べていくことで行う。このとき実施の形態の項で説明した所定の区切り記号を用いる。例えば、図53(a)に示す化合物の原子番号7をフォーカスとする場合、図53(b)に示すようなツリーから同図に示すようなHYPER Codeが生成される。

【0143】以上本発明の実施例について説明したが、本発明は以上の実施例にのみ限定されずその他各種の付加変更が可能である。例えば、炭素のノードコードが、フォーカスノード以外でコードが2になる場合、次スフィアのノードとのボンドが二重結合のときは2に代えて6を使用するようにしても良い。これは、ノードコード2はCHを表現しているが、次のようにケミカルシフト値がその環境によって異なっている。

-CH= 5.0 ppm~10.0 ppm

10 -CH< 2.5 ppm~7.5 ppm

従って、この情報を反映させるため、次スフィアと二重結合しているときは6に変更する。一例を図54に示す。これにより、次のスフィアが不明でも、2と6により異なった構造であることが容易に判明する。

【0144】

【発明の効果】以上説明したように本発明によれば以下のような効果を得ることができる。

【0145】炭素のノードに対してその結合水素数に応じて異なるノードコードを割り当てるようにした為、炭素の結合水素数がノードコードから直ちに判明し、また最終スフィアに炭素が存在する場合でも結合水素数の情報が失われない。よって、水素の結合する元素の環境を精度良く表現することができる。

【0146】炭素に加えて、窒素および酸素のノードに対してもその結合水素数に応じて異なるノードコードを割り当てる構成では、水素の結合する元素の環境をより一層精度良く表現することができる。

【0147】1つの部分構造のコード化において複数の閉環コードを使用する必要がある場合、予め定められたそれぞれ異なる複数の閉環コードから循環的に1つずつ閉環コードを選択して使用するようにした為、1つの部分構造に複数の環が存在する場合にその閉環位置を正確にコード化できる。よって、コードから構造式に組み立てるときに正しい構造の再現ができる。

【0148】有機化合物を構成する元素の種類および各元素間の結合状態を記述した入力ファイルから自動的に有機化合物の部分構造のコードを生成することができる。

【図面の簡単な説明】

40 【図1】本明細書で使用する用語の説明のために用いる有機化合物の構造式の例を示す図である。

【図2】ノードコードの変換規則表の例を示す図である。

【図3】ボンドコード表の一例を示す図である。

【図4】閉環コード表の一例を示す図である。

【図5】HYPER Codeの一般的なフォーマットを示す図である。

【図6】優先順位表の一例を示す図である。

50 【図7】有機化合物の構造式の例とそのHYPER Codeを示す図である。

【図8】有機化合物の構造式の他の例とそのHYPER Codeを示す図である。

【図9】HYPER Code自動生成装置の一実施例のブロック図である。

【図10】入力ファイルである結合表の一例を示す図である。

【図11】図10の結合表で示される有機化合物の構造式を示す図である。

【図12】HYPER Code自動生成装置の機能ブロック図である。

【図13】コード生成機構の処理の概要を示すフローチャートである。

【図14】図13のステップS4の詳細を示すフローチャートである。

【図15】木構造を作成する際に使用するセルの構成例を示す図である。

【図16】作成された木構造の一例を示す図である。

【図17】図13のステップS5の詳細を示すフローチャートである。

【図18】正規化された木構造の一例を示す図である。

【図19】図13のステップS6の詳細を示すフローチャートである。

【図20】出力ファイルに出力されたHYPER Codeとその付加情報の例を示す図である。

【図21】本発明の別の実施例における処理の流れを示すフローチャートである。

【図22】有機化合物（ナフタレン誘導体）の構造とその初期ツリーを示す図である。

【図23】ツリーをコンピュータ内で保存するための表（ノードテーブル構造体）の構成図である。

【図24】シンボルの記述順序の説明図である。

【図25】偶数環の存在する多環式化合物の例を示す図である。

【図26】図25に示す多環式化合物から生成されるツリーの例を示す図である。

【図27】図25に示す多環式化合物から生成される別のツリーの例を示す図である。

【図28】ノード α 、 β 配下に偶数閉環があり、 α 、 β が同一シンボルで、シンボルを見ただけではどちらを先に記述するか分からないツリーの例を示す図である。

【図29】ノード α 、 β 配下に偶数閉環があり、 α 、 β が同一シンボルで、シンボルを見ただけではどちらを先に記述するか分からないツリーに対する正規化ルートの説明図である。

【図30】対称性の高い構造例を示す図である。

【図31】正規化アルゴリズムを考える上での基本事項の説明図である。

【図32】正規化アルゴリズムの適用例の説明図である。

【図33】正規化アルゴリズムの適用例の説明図であ

る。

【図34】正規化アルゴリズムの適用例の説明図である。

【図35】ツリーの正規化処理の全体の概略フローチャートである。

【図36】親ノードX配下の並べ換え処理の詳細なフローチャートである。

【図37】2ノードA、B配下の入れ替え処理の詳細なフローチャートである。

10 【図38】DT順位に従ってツリーを並べ替える処理の詳細なフローチャートである。

【図39】別ツリーを作成して優先度の高いツリーを取得する処理の詳細なフローチャートである。

【図40】親ノードX配下のノード並べ替え処理のアルゴリズムスケルトンを示す図である。

【図41】2ノードA、B配下の入れ替え処理のアルゴリズムスケルトンを示す図である。

【図42】DT順位に従ってツリーを並べ替える処理のアルゴリズムスケルトンを示す図である。

20 【図43】別ツリーを作成して優先度の高いツリーを取得する処理のアルゴリズムスケルトンを示す図である。

【図44】図39のステップS508で行われる、DT1、DT2の各左側より、DDを作成する処理のイメージを示す図である。

【図45】図39のステップS509で行われるDD内の左右の順位決定処理のイメージを示す図である。

【図46】閉環シンボルの並べ替え処理の説明図である。

30 【図47】閉環シンボルの並べ替え処理の説明図である。

【図48】環構造を含む有機化合物の場合に同じ構造から異なるHYPER Codeが生成される可能性があることの説明図である。

【図49】閉環シンボル対の説明図である。

【図50】環構造を含む有機化合物における閉環シンボルの並べ替え処理の具体例を示す図である。

【図51】環構造を含む有機化合物における閉環シンボルの並べ替え処理の基本的な考え方の説明図である。

40 【図52】環構造を含む有機化合物における閉環シンボルの並べ替え処理の具体例を示す図である。

【図53】ツリーからHYPER Codeへの変換処理の説明図である。

【図54】結合状態によりノードコード2に代えてノードコード6を使用する例の説明図である。

【図55】従来のHOSE Codeの説明図である。

【符号の説明】

100…中央処理装置

101…メモリ

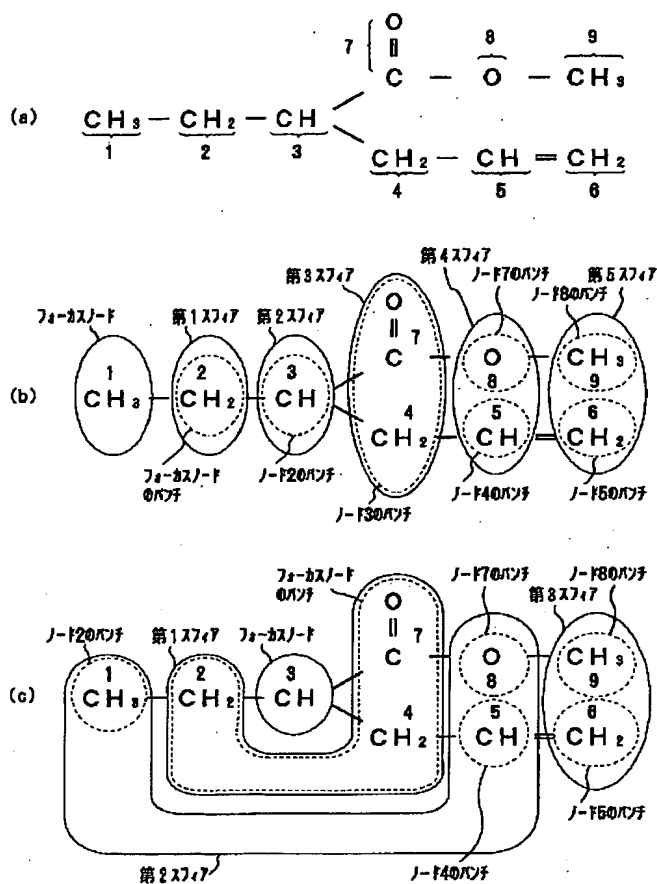
102…プログラム記録媒体

50 103…入力ファイル

104…出力ファイル
 105…入力装置
 106…表示装置
 107…プリンタ
 200…ヘッダ
 201…カウントライン
 202…アトムブロック
 203…ボンドブロック
 204…シフトブロック
 300…コード生成機構
 301…パラメータ入力手段
 302…パラメータ格納領域
 303…入力ファイル読み込み手段
 304…入力ファイル格納エリア

* 305…木構造生成手段
 306…木構造生成エリア
 307…木構造正規化手段
 308…コード化手段
 309…出力手段
 310…コード格納エリア
 311…変換規則表
 312…ボンドコード表
 313…閉環コード表
 10 314…優先順位表
 F…フォーカス情報
 SP1～SPN…第1スフィア情報～第Nスフィア情報
 BC…ボンドコード
 * NC…ノードコード

【図1】



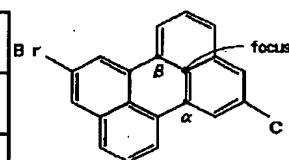
【図4】

閉環コード	&, !, @, #, ?, ~ (チルド)
-------	------------------------

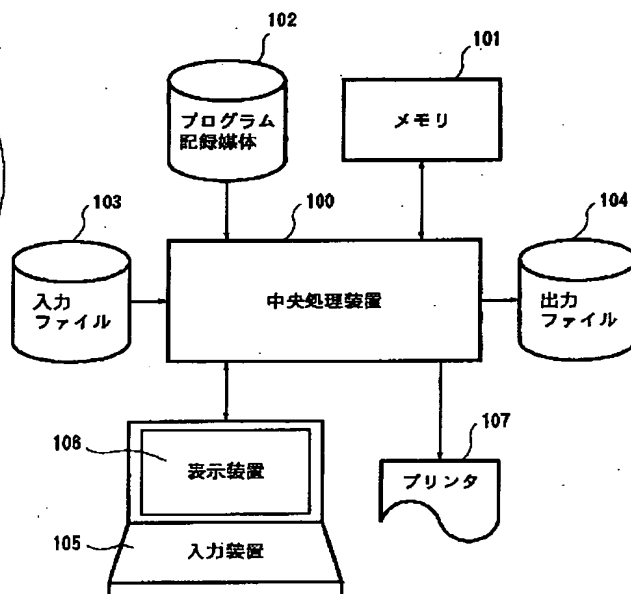
【図3】

ボンド	ボンドコード
単結合	無し
二重結合	=
三重結合	%
芳香族結合	*
その他の結合	>

【図25】



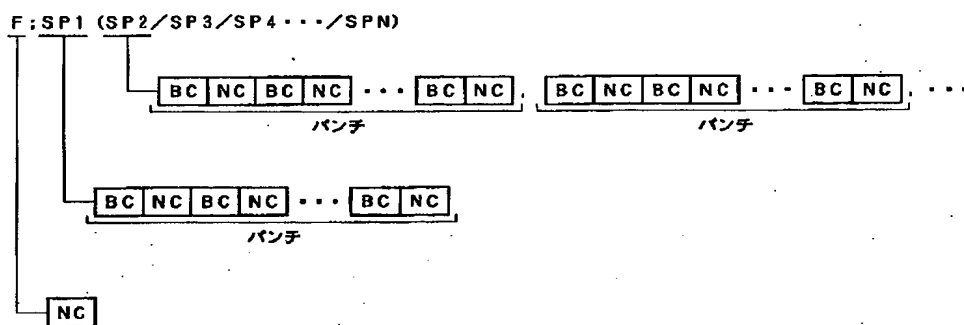
【図9】



【図2】

元素記号	ノードコード	
C	結合水素数0 (C)	1
	結合水素数1 (CH)	2
	結合水素数2 (CH ₂)	3
	結合水素数3 (CH ₃)	4
N	結合水素数0 (N)	N
	結合水素数1 (NH)	E
	結合水素数2 (NH ₂)	G
	結合水素数3 (NH ₃)	J
O	結合水素数0 (O)	O
	結合水素数1 (OH)	L
Si		Q
Cl		X
Br		Y
I		" I "
Y		" Y "
1, Y以外の1文字の元素記号 元素記号と同じ (例) B, F, K, P, S, U, V, W		
2文字の元素記号 (M ₂) " MM" (例) Fe→" FE", Cu→" CU"		
その他、水素は一般にコード化しないが、必要なときは" H", 重水素はDで表す		

【図5】



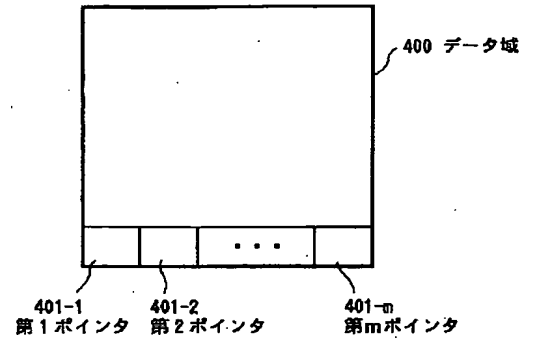
【図6】

優先順位	高 ←	→ 低
コード		
ボンドコード	%, =, *, 無し (単結合)	
ノードコード	1, 2, 8, 3, 4, O, L, N, E, G, J, S, P, Q, F, X, Y, " I ", " Y ", " MM", 隠蔽シンボル (&, !, @, #, ?, _ , ~ (ほか)) (MMは原子番号取)	

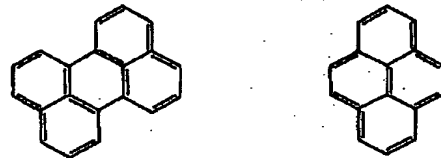
【図10】

ETHYL (4-METHOXYPHENYL)ACETATE									
PNS2802									
14	14	0	0	0	0	0	0	0	12
-1.3300	2.1900	0.0000	C	0	0	0	0	0	0
-1.3300	3.1200	0.0000	O	0	0	0	0	0	0
-2.2200	2.4800	0.0000	C	0	0	0	0	0	0
-2.2300	1.4900	0.0000	C	0	0	0	0	0	0
-2.2300	-0.3900	0.0000	C	0	0	0	0	0	0
-0.4800	2.5400	0.0000	O	0	0	0	0	0	0
-3.0400	1.0200	0.0000	C	0	0	0	0	0	0
-1.4200	1.0200	0.0000	C	0	0	0	0	0	0
-1.4200	0.0800	0.0000	C	0	0	0	0	0	0
-3.0400	0.0800	0.0000	C	0	0	0	0	0	0
-2.2200	-1.2900	0.0000	O	0	0	0	0	0	0
0.4400	2.2100	0.0000	C	0	0	0	0	0	0
-1.2200	-1.2900	0.0000	C	0	0	0	0	0	0
1.2200	2.5800	0.0000	C	0	0	0	0	0	0
2	1	2	0	2	0	0			
3	1	1	0	2	0	0			
4	3	1	0	2	0	0			
5	9	4	0	1	0	0			
6	1	1	0	2	0	0			
7	4	4	0	1	0	0			
8	4	4	0	1	0	0			
9	7	4	0	1	0	0			
10	8	4	0	1	0	0			
11	5	1	0	2	0	0			
12	6	1	0	2	0	0			
13	11	1	0	2	0	0			
14	12	1	0	2	0	0			
10	5	4	0	1	0	0			
N	HSF	1	3	0	3.5400	3.5400	0.0000	0	
N	HSF	1	7	0	7.2000	7.2000	0.0000	0	
N	HSF	1	8	0	7.2000	7.2000	0.0000	0	
N	HSF	1	9	0	6.8600	6.8600	0.0000	0	
N	HSF	1	10	0	6.8600	6.8600	0.0000	0	
N	HSF	1	12	0	4.1300	4.1300	0.0000	0	
N	HSF	1	13	0	3.7900	3.7900	0.0000	0	
N	HSF	1	14	0	1.2200	1.2200	0.0000	0	
N	HSC	1			90.0000	0.0000			
N	SLV	1			CDCL3				
N	REF	1			PNS2802				
N	END								

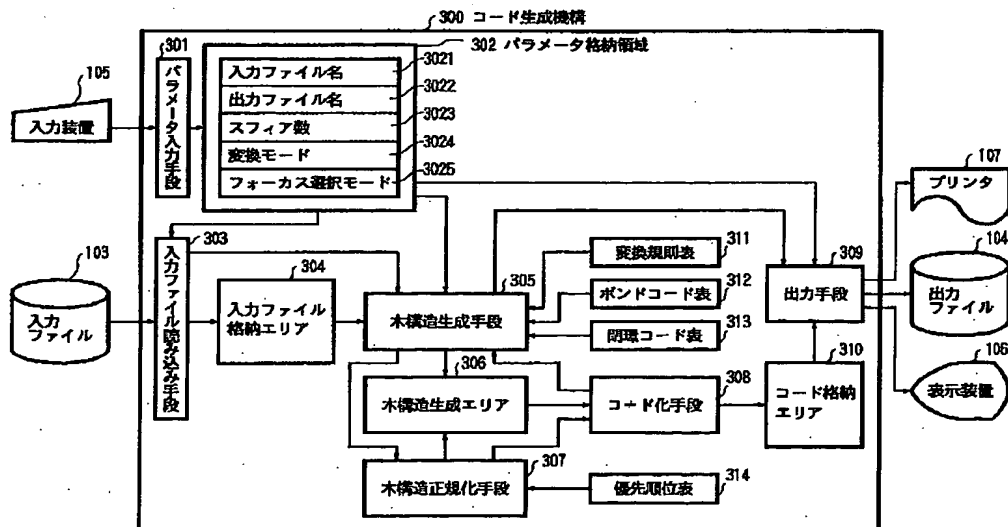
【図15】



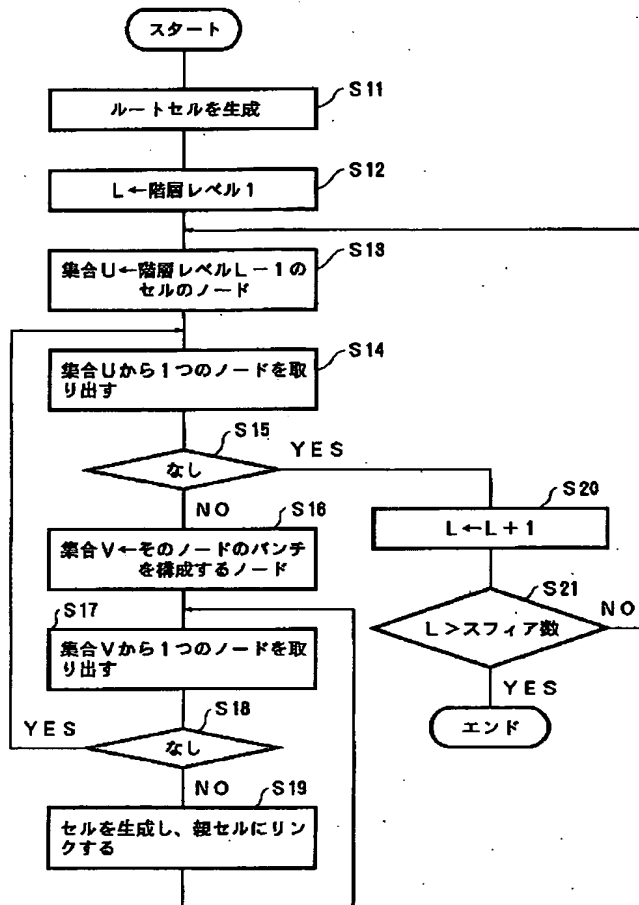
【図30】



【図12】



【図14】

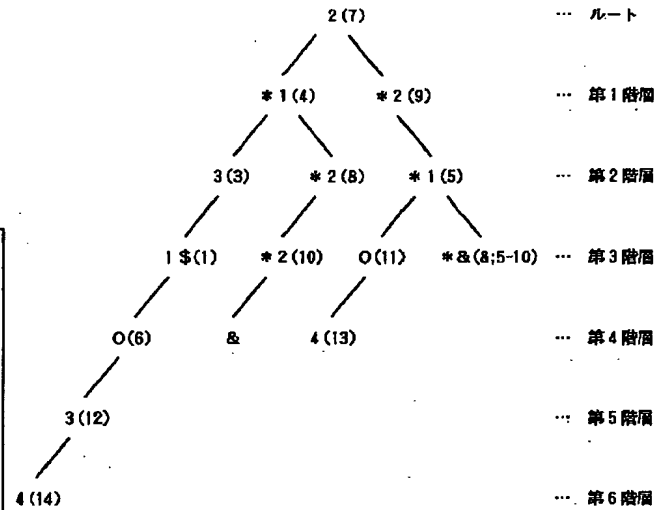


【図20】

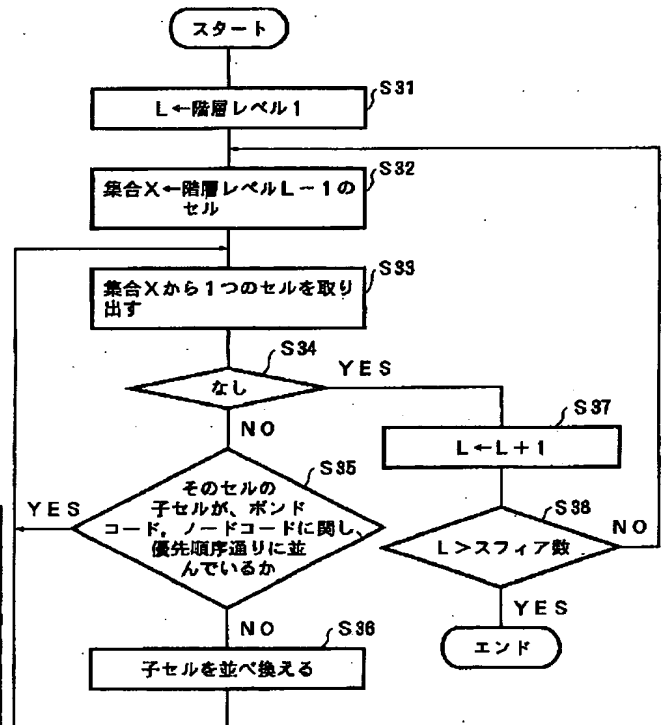
104

501	3CD 90h	3.540	3;1\$1(0, *2*2/3, *2, *2/4, *1, *8/, *80/4)
502	7CD 90h	7.200	2; *1*2(*23, *1/*2, 1\$, *80/*8, 0, 4/3, /4)
503	8CD 90h	7.200	2; *1*2(*23, *1/*2, 1\$, *80/*8, 0, 4/3, /4)
504	9CD 90h	6.880	2; *1*2(*20, *1/*2, 4, *83/*8, . 1\$/0/3)
505	10CD 90h	6.880	2; *1*2(*20, *1/*2, 4, *83/*8, . 1\$/0/3)
506	12CD 90h	4.130	3; 40(. 1\$/3/1/*2*2/*2, *2)
507	13CD 90h	3.790	4; 0(1/*2*2/*2, *2/*1, *8/*83)
508	14CD 90h	1.220	4; 8(0/1\$/3/1/*2*2)

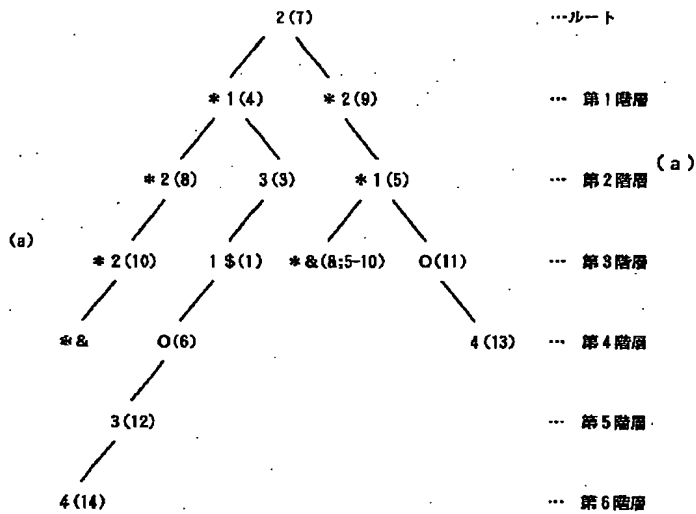
【図16】



【図17】

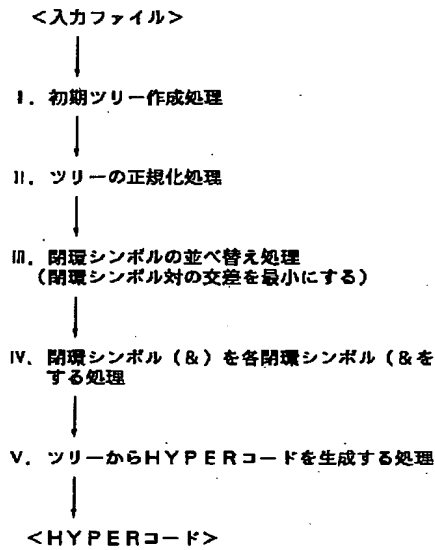


【図18】

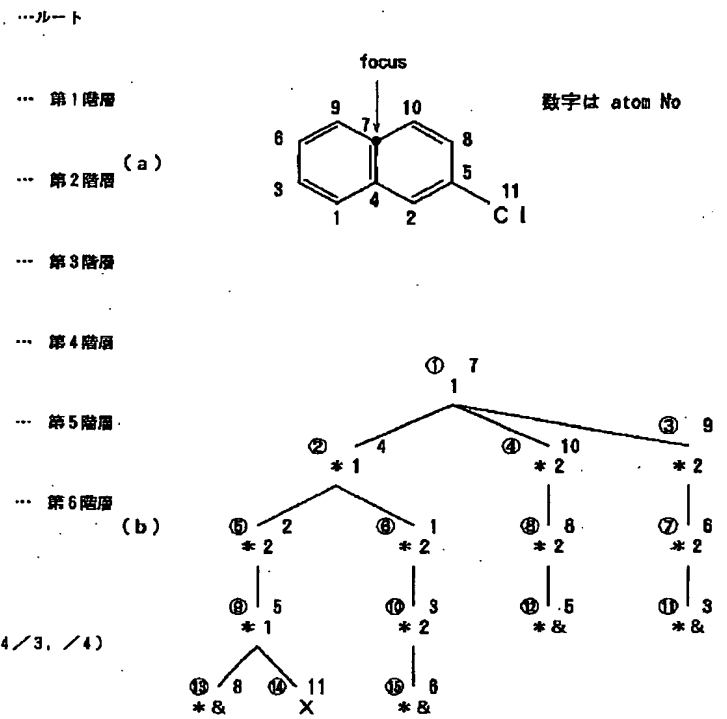


(b) 2; *1*2 (*2 3, *1/*2, 1\$, *&O/*&, O, 4/3, /4)

【図21】



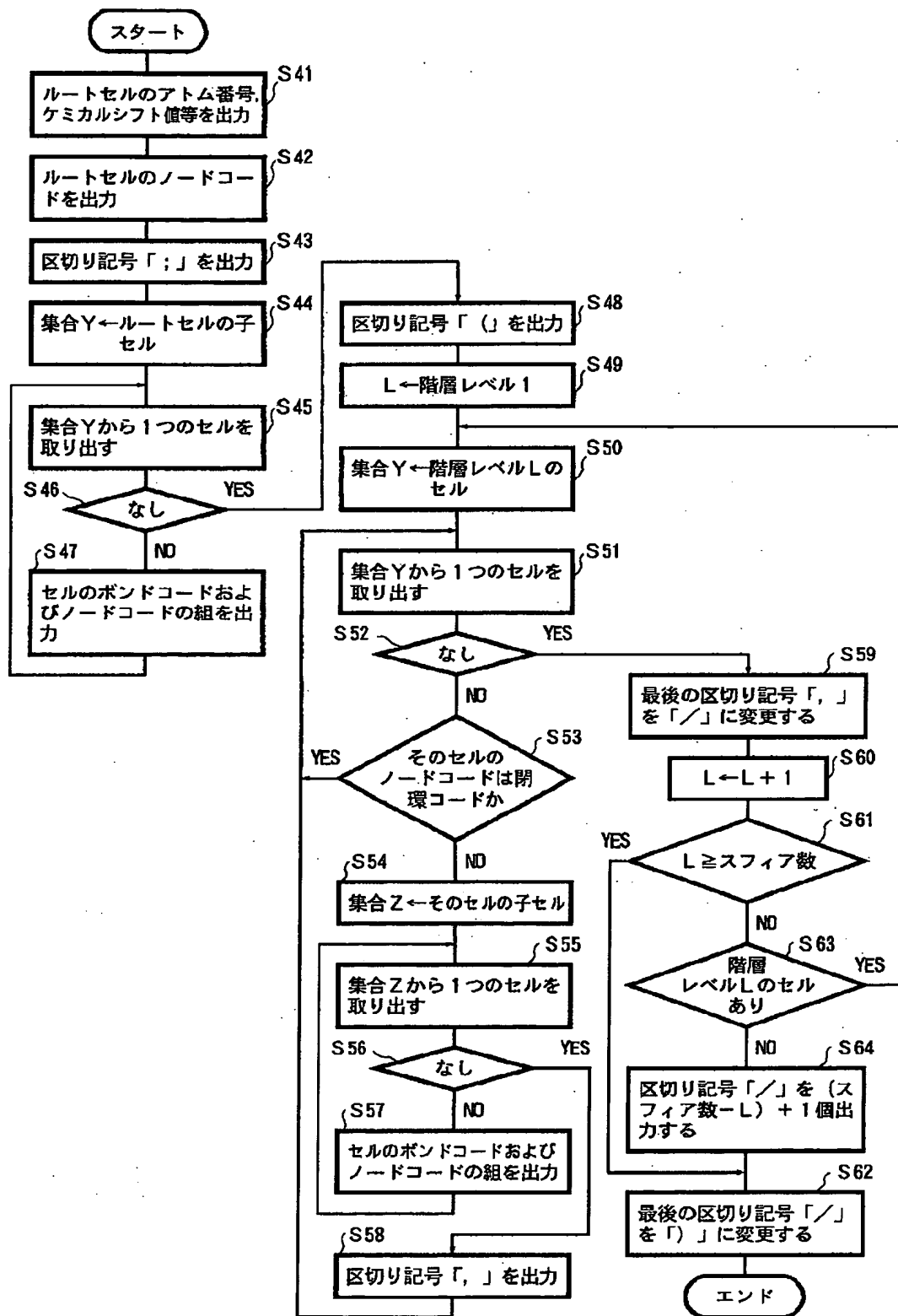
【図22】



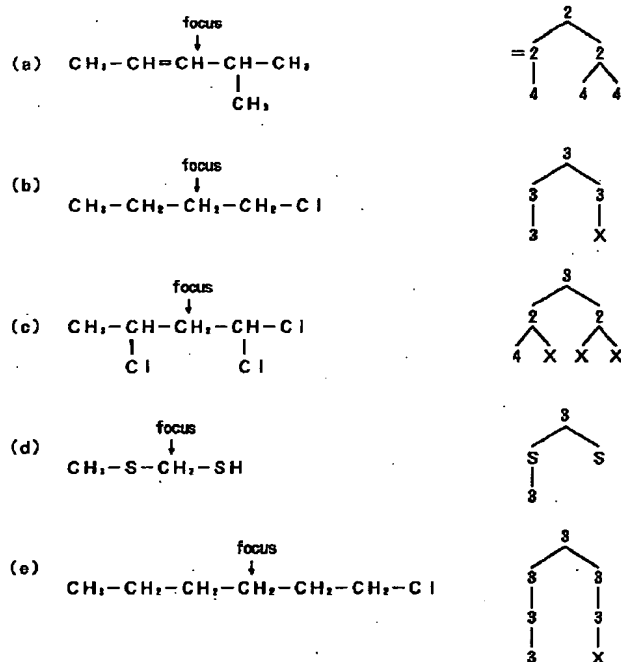
【図23】

ノードID	atomNo	シンボル	閉環相手ノードID	親ノードID	子ノードID (左端のノードから記述)
1	7	1	0	0	2 3 4
2	4	* 1	0	1	5 6
3	9	* 2	0	1	7
4	10	* 2	0	1	8
5	2	* 2	0	2	9
6	1	* 2	0	2	10
7	6	* 2	0	3	11
8	8	* 2	0	4	12
9	5	* 1	0	5	13 14
10	3	* 2	0	6	15
11	3	* &	15	7	0
12	5	* &	13	8	0
13	8	* &	12	9	0
14	11	X	0	9	0
15	6	&	11	10	0
0	0	0	0	0	0
.
.
.

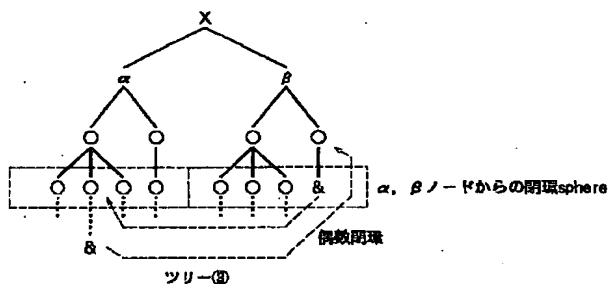
【図19】



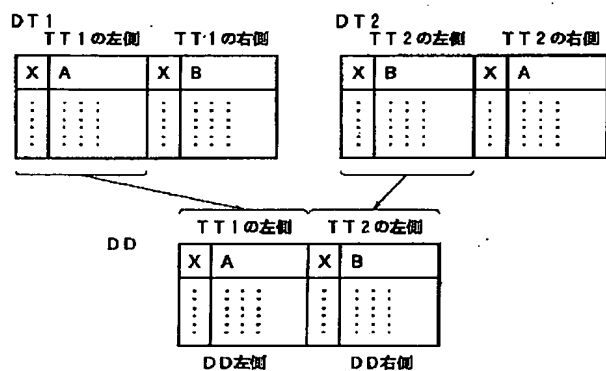
【図24】



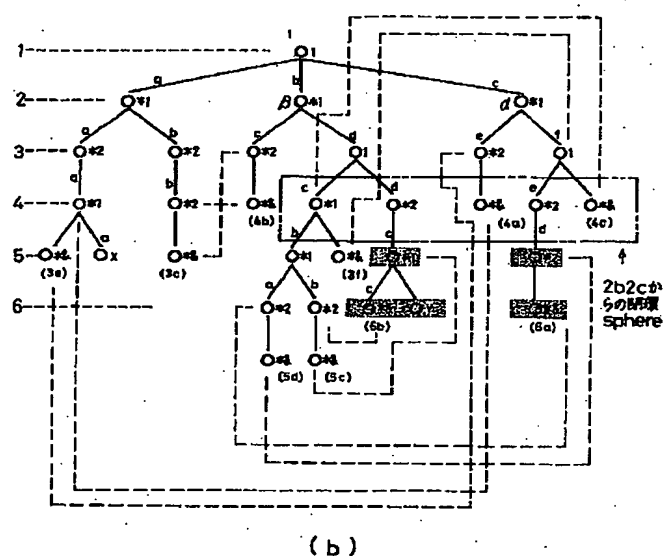
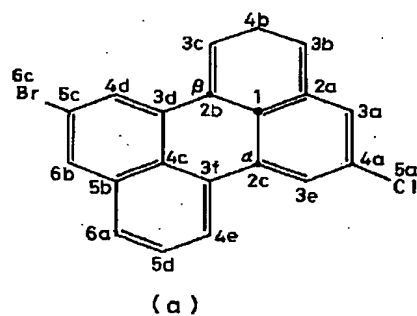
【図28】



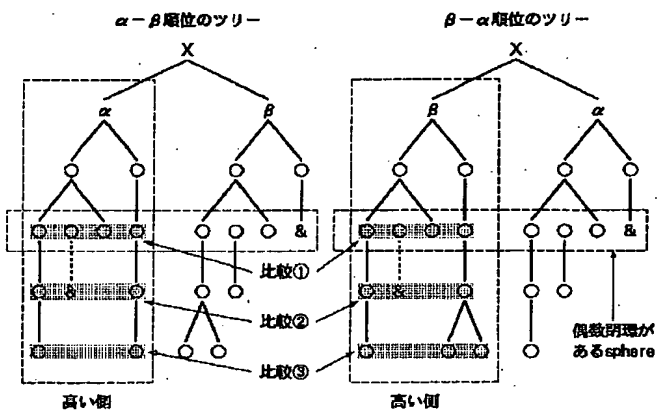
【図44】



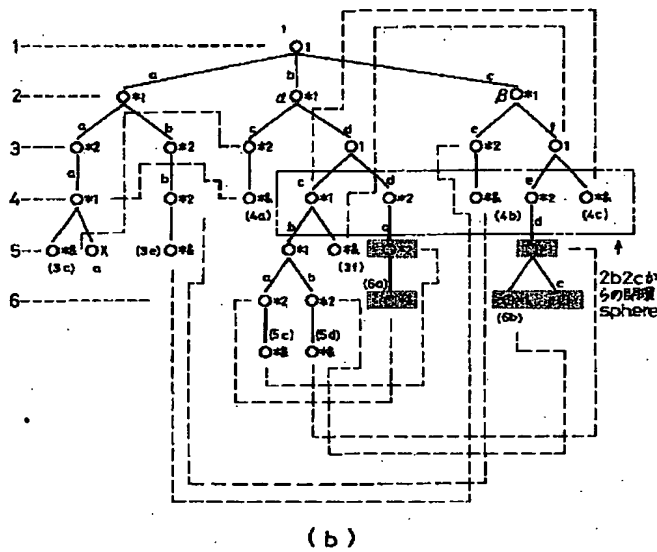
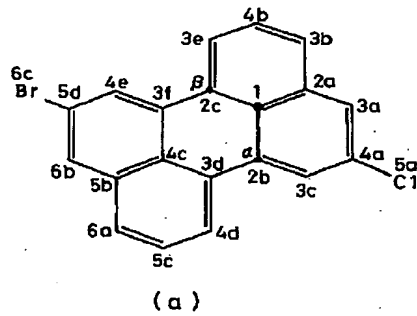
【図27】



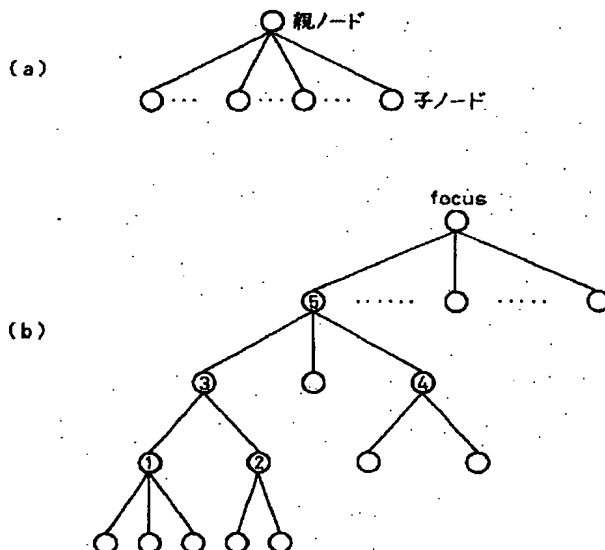
【図29】



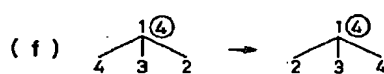
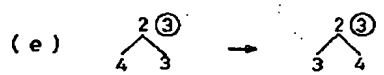
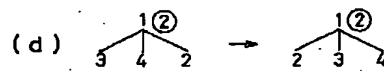
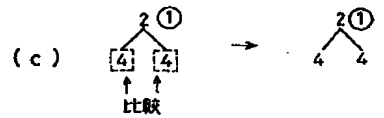
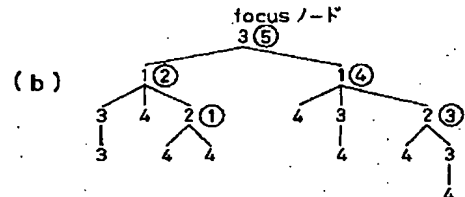
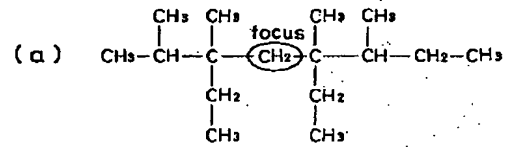
【図26】



【図31】



【図32】



【図41】

2ノード配下の入れ替え処理 (DT, TT) {

DTの順位を決定する。

if (DT順位が決定できたなら) {
DTの順位に従ってツリー (TT) を並べ替える。
}

if (親ノードsphereが最終sphereより小さく
かつ DT順位が未定なら) {

DTに次sphereを追加 (DT, TT) ;

親ノードsphere ← 親ノードsphere + 1

2ノード (A, B) 配下の入れ替え処理 (DT, TT) ;

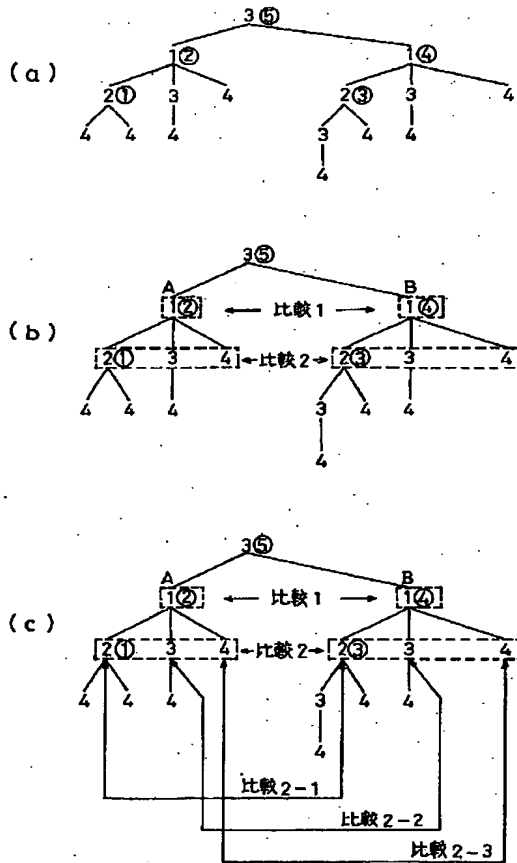
else {
if (順位決定sphereに偶数個のノードがある) {

別ツリーを作成して優先度の高いツリーを取得 (DT, TT) ;

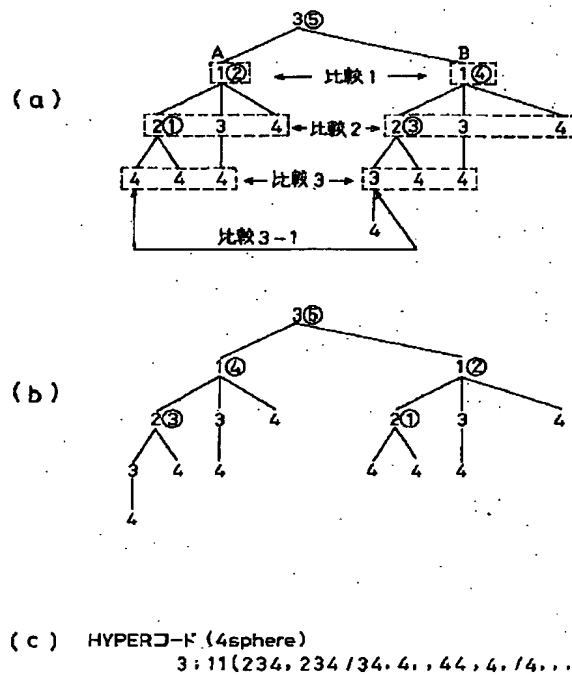
}

}

【図33】

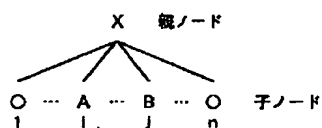


【図34】



【図40】

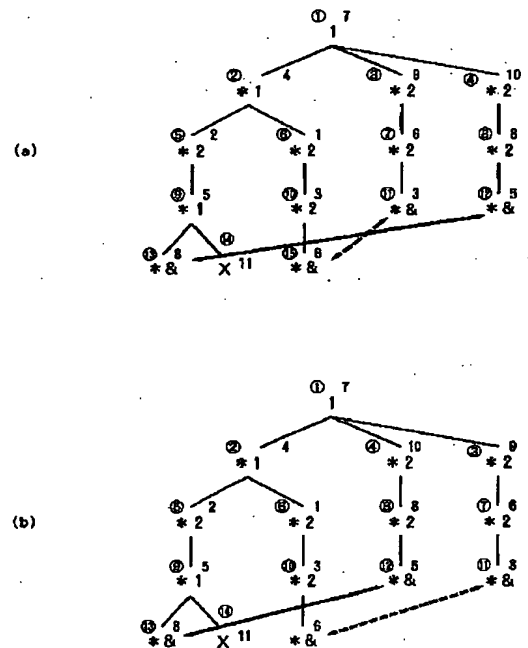
ノードX配下の並べ替え処理(親ノードX、TT) {
 while (親ノードXに子ノードがある間) {
 cldid ← Xの子ノード;
 親ノードsphere ← 親ノードsphere + 1
 ノードX配下の並べ替え処理(cldid, TT);
 }
 for (i=1; i<=n; i++) {
 for (j=i+1; j<=n; j++) {
 DTを作成して、親ノードXとi, j番目の子ノード(A, B)をセット;
 2ノード(A, B)配下の入れ替え処理(DT, TT);
 }
 }
}



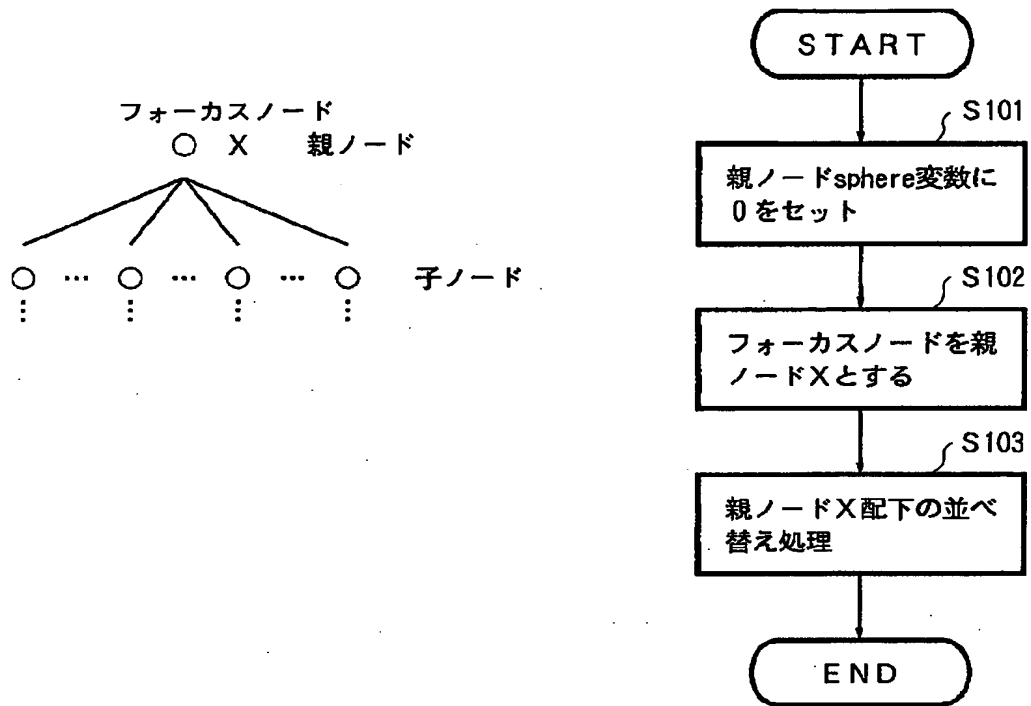
左側		右側	
親ノード	子ノード	親ノード	子ノード
X	A	X	B

DT構造体の初期状態

【図50】



【図35】



【図42】

DT順位に従ってツリーを並べ替える (DT, TT) {

if (DT順位が右制優先で決定しているならば) {
DTの順位に従ってTTを並べ替える。
}

ツリー上のAとB、及びそれらの配下ノードを入れ替える。

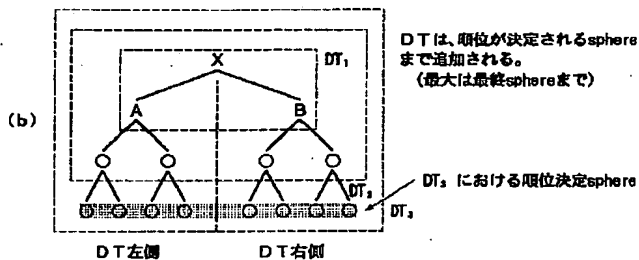
(a) ツリー全体の閉環ノードを入れ替える。

X ← DT内の最上位の親ノード
while (Xに子ノード(cldid)がある間) {

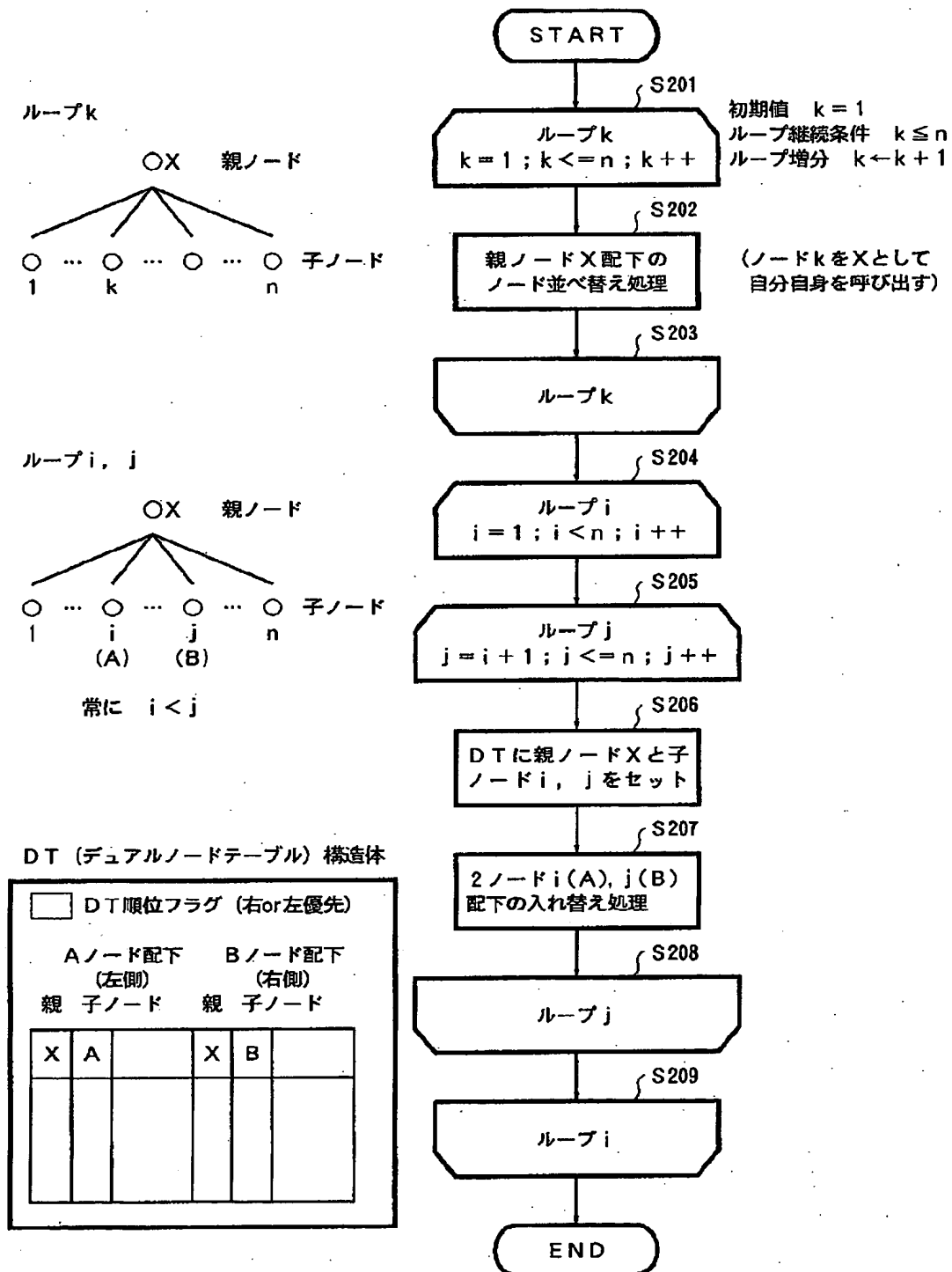
親ノードsphere ← 親ノードsphere + 1

ノードX配下の並べ替え処理 (cldid, TT);

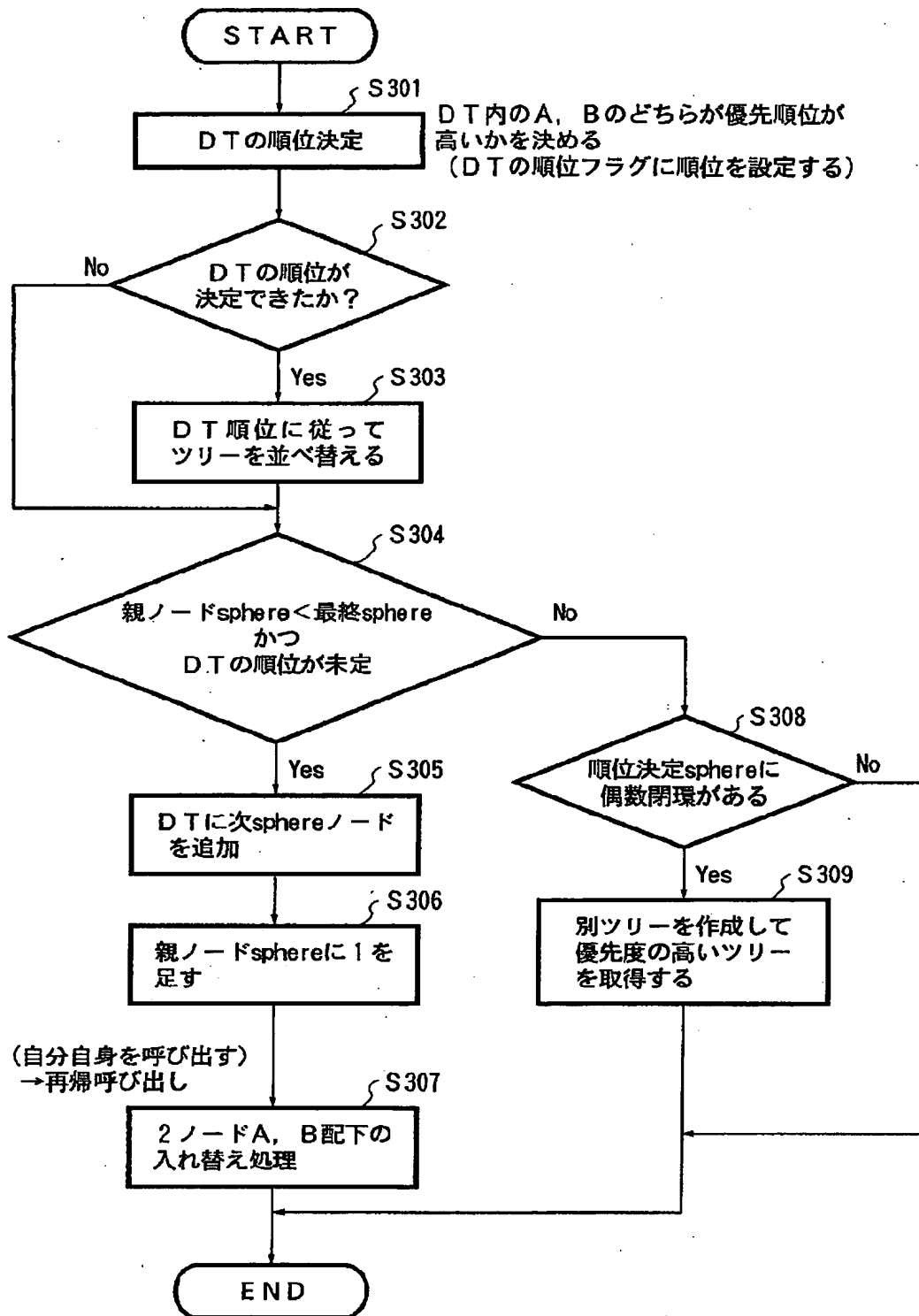
}



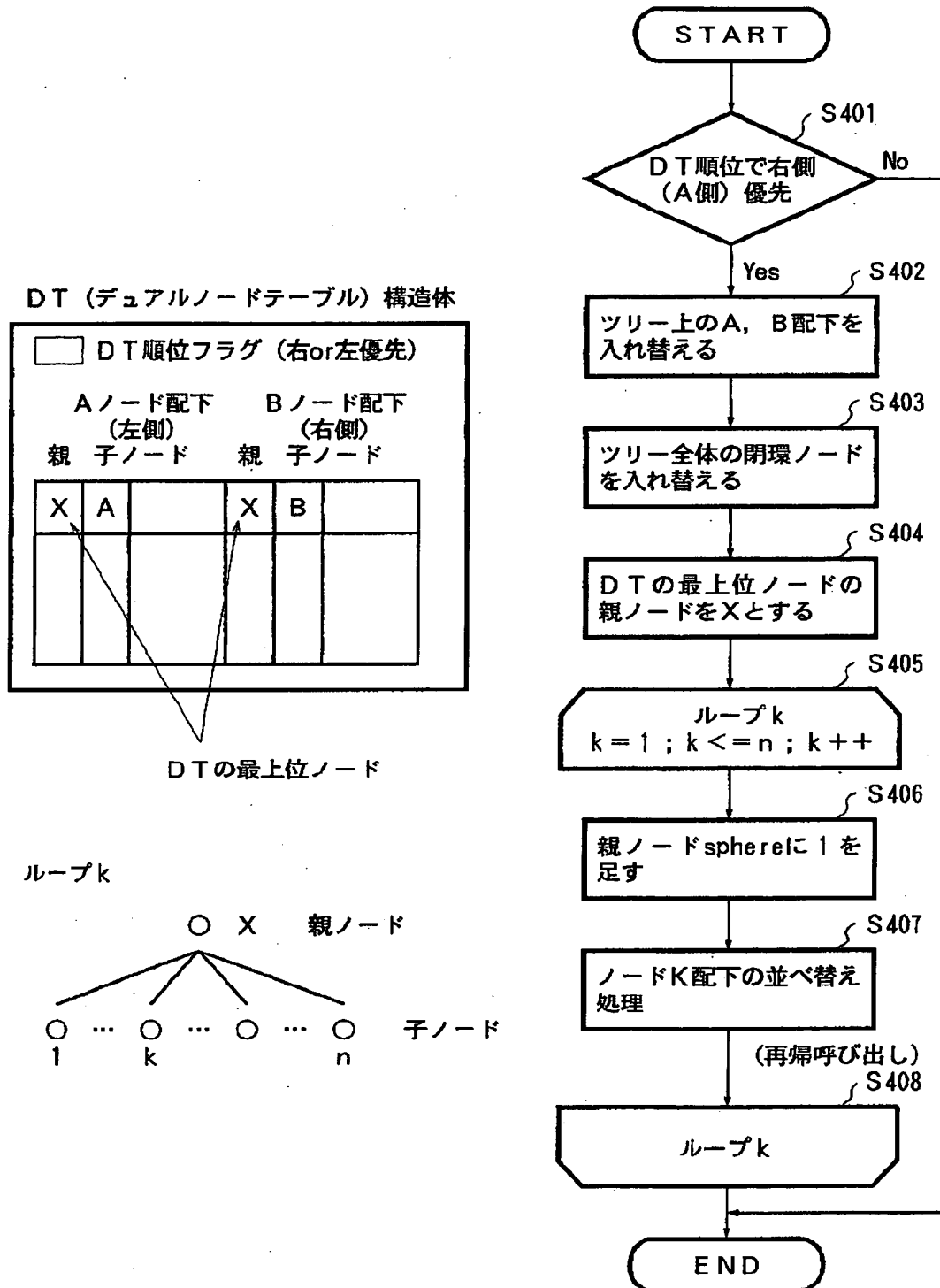
【図36】



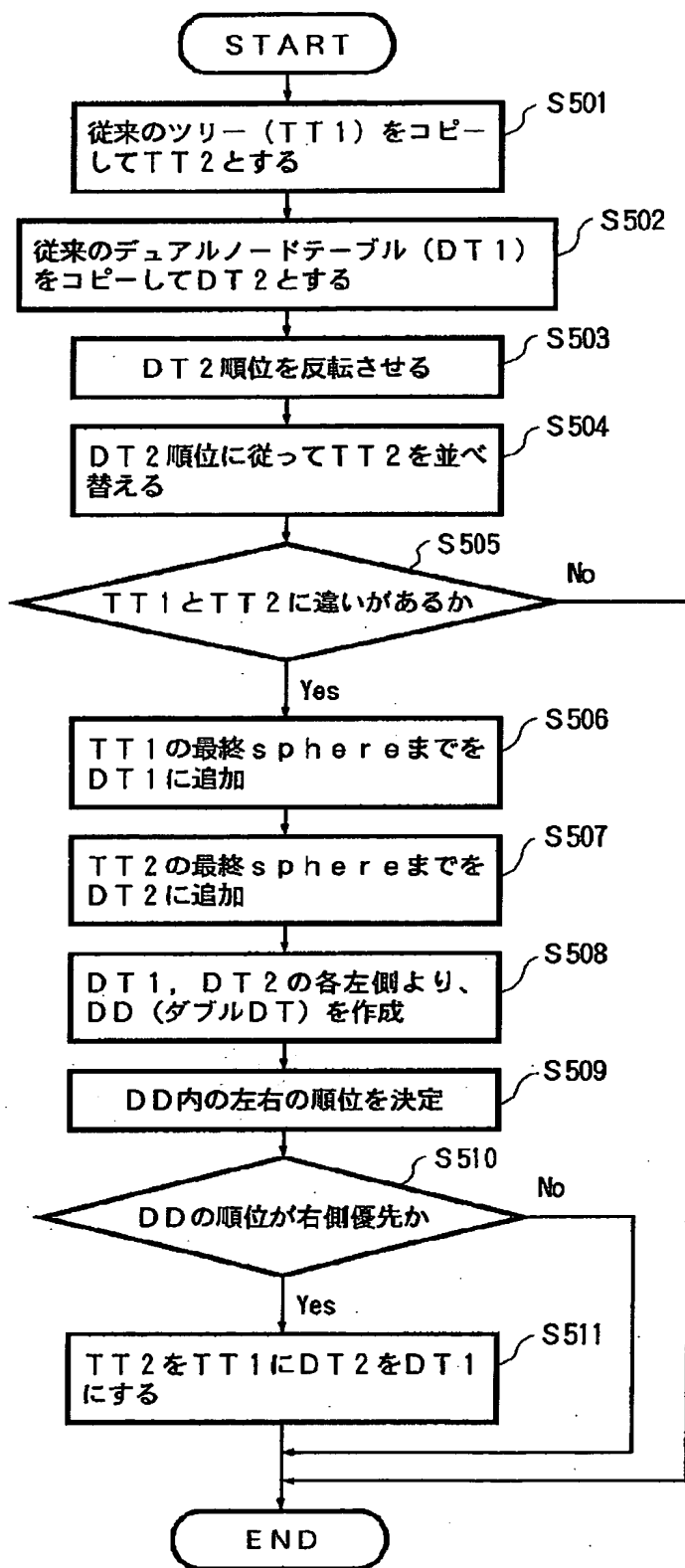
【図 37】



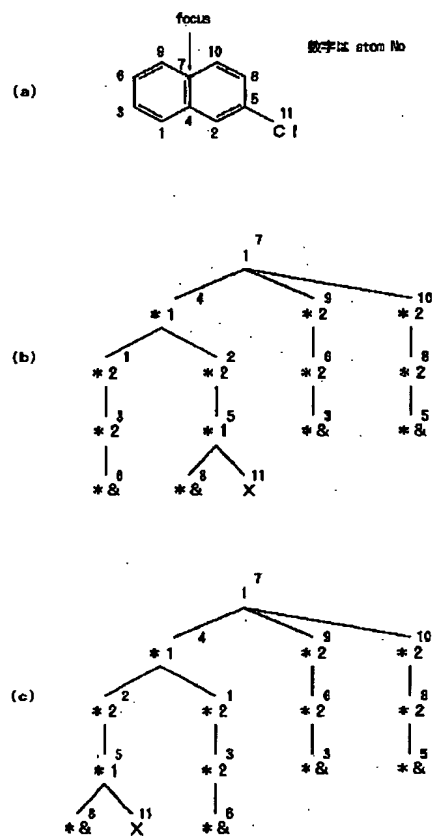
【図38】



【図39】

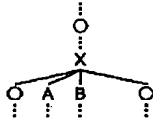


【図46】

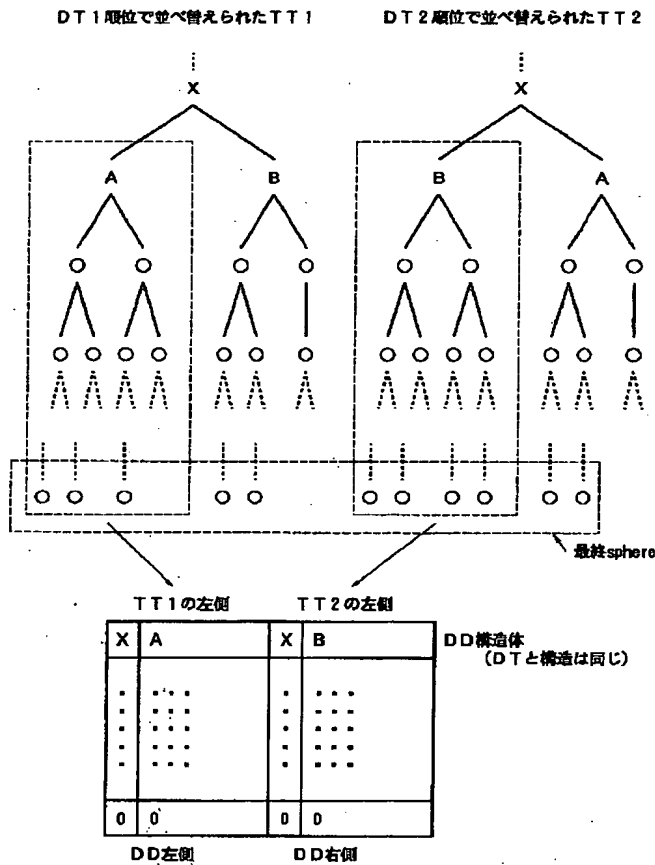


【図43】

別ツリーを作成して優先度の高いツリーを取得 (DT1, TT1) {
 DT1をDT2に、TT1をTT2にコピーする
 DT2の順位を反転。
 DT2の順位に従ってツリーTT2を並べ替える (DT2, TT2) ;
 if (TT1とTT2に違いがあるなら) {
 TT1の最終sphereまでをDT1に追加
 TT2の最終sphereまでをDT2に追加
 DT1、DT2の各左側よりDDを作成;
 DD内の左右の順位を決定する。
 if (DD順位が右側優先) {
 TT2をTT1、DT2をDT1とする
 }
 }
}

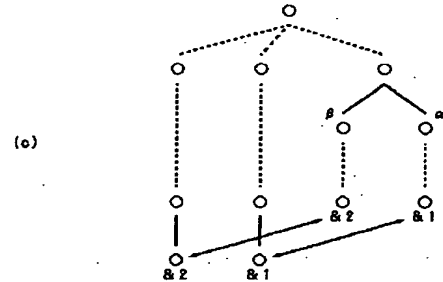
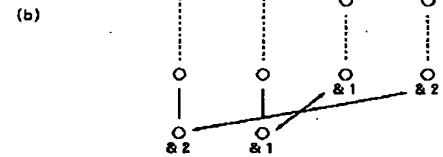
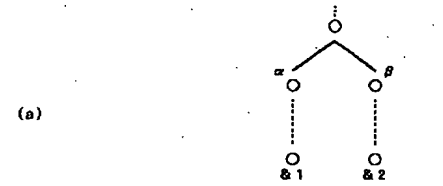


【図45】

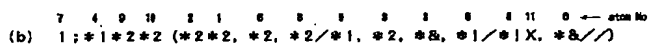
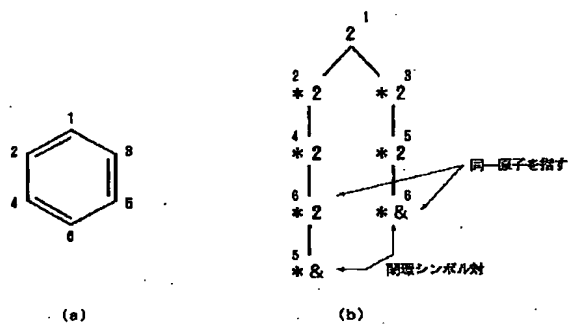


DTの左右順位決定と同じルール (順位決定関数) でDDの左右順位を決める。

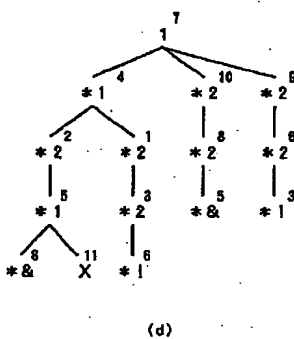
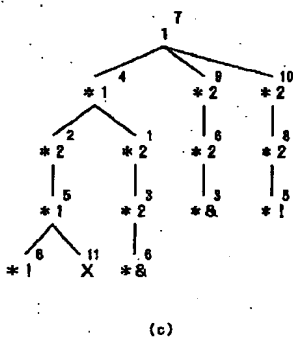
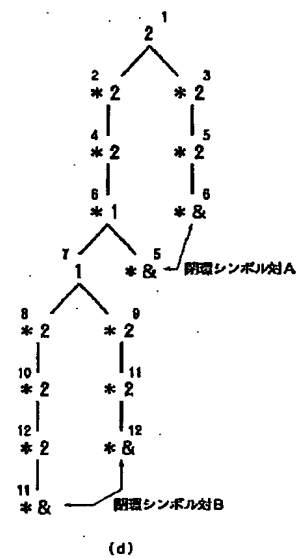
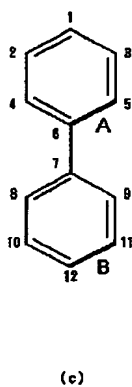
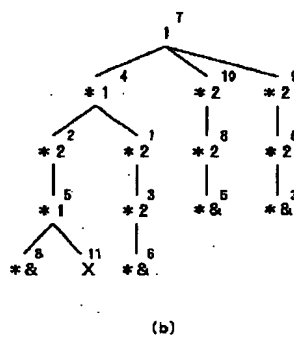
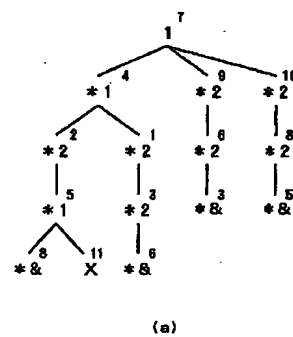
【図51】



【図 49】



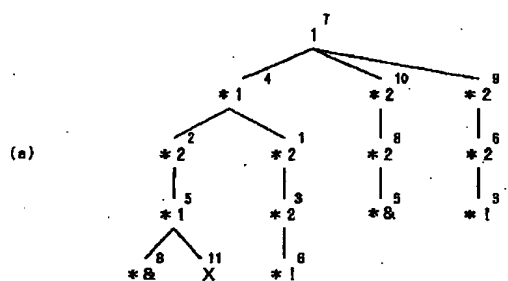
【図 48】



1:014202(+202, 02, +2/01, +2, 00, +1/01X, 00//)

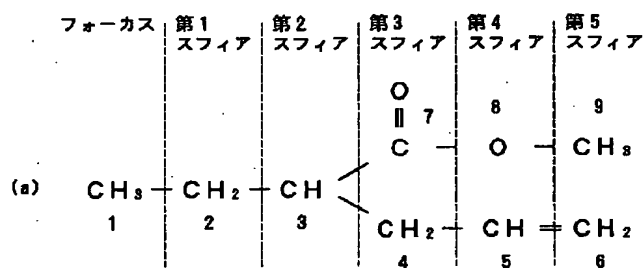
(f)

【図52】

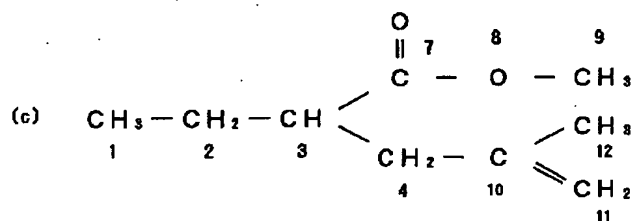


(b) 7 4 10 9 2 1 8 6 5 3 6 3 8 11 6 — atom No
1: *1*2*2 (*2*2, *2, *2/*1, *2, *8, *1/*8X, *1//)

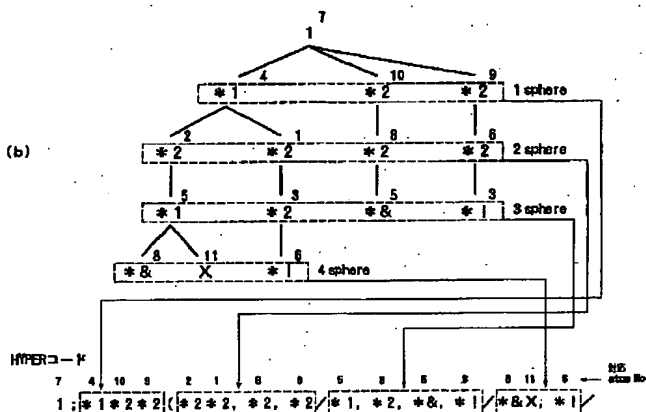
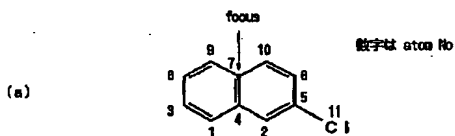
【図55】



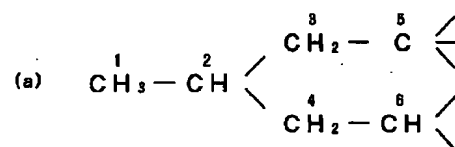
(b) C (C/CSC/O, C)



【図53】



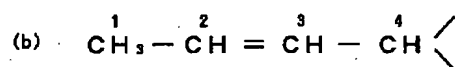
【図54】



HYPER Code ケミカルシフト値 (フォーカス)

フォーカスノード1 4 ; 2 (3 3 / ... 0.91 ppm

フォーカスノード2 2 ; 3 3 4 (1, 2, / ... 1.54 ppm



HYPER Code ケミカルシフト値 (フォーカス)

フォーカスノード1 4 ; 8 (= 2 / ... 1.79 ppm

フォーカスノード2 2 ; = 2 4 (2, / ... 5.60 ppm